Chapter 4

MODELING THE FLOWS

This chapter examines the methodologies used in estimating coal, non-metallic mineral, farm products, and manufacturing commodity flows for the state of Indiana. We will begin with an overview of the research design used here. This will be followed by specific discussions of the procedures used in the commodity production and attraction phases of the traffic generation analysis. The distribution of traffic, identifying the origins and destinations of commodity flow shipments, is discussed next. Once the origins and destinations of flows are known the next concern is estimating the modes taken by these flows; the method of splitting this distributed traffic between modes is discussed next. The final step in the analysis is to assign the traffic to the appropriate transport networks. In this case these are the highway and railroad networks discussed earlier in this report.

Comments on Research Design

The brief discussion above outlines the major steps in the planning process that were undertaken here. It appears to be quite similar to the urban transportation planning process, and it is. It is different in that studies of this scale are not undertaken very often. This has implications for the analysis that separates it from comparable studies of urban areas. For example, trip production and attraction in the urban transportation planning process is almost routine and standards exist for identifying these values for families of different size or with different levels of automobile ownership. In the regional commodity flow case examined here such standards do not exist and they had to be developed here; they were developed for urban transport studies nearly a quarter century ago.

Similarly the distribution of traffic is reasonably well understood in urban transport planning and model parameters (e.g., friction factors) for the movement of people are known to some extent. The distribution of goods is not so well understood and little help is available in this area from urban transport planning since the latter often uses a constant proportion to represent movement of goods or commodities. This can be an advantage. In the present case it has resulted in the use of a flow model that offers much better estimates than the traditional approaches one would encounter in urban transport studies.

Modal split of commodity traffic is an area of few generalizations in the area of regional transport analysis. It is generally recognized that bulk goods are moved by rail if distances are in excess of 700 miles and by trucks if the mileages are less than this. However, the primary, though not exclusive focus here, is manufactured goods and these bulk commodity generalizations are weak in this context. There are mathematical models available that could be used to determine such modal splits, but it is not clear that such models offer any real advantages over current patterns of modal traffic. For example, the criteria used by shippers are not necessarily the same for motor carriers and rail carriers. Motor carriers tend to cost more than rail carriers, but are faster than rail as well. In some cases it is not clear that these modes are competitive on a cost or time basis. In any event, this project uses 1993 patterns based on the Commodity Flow Survey of that year [1]. The procedure used recognizes the assignment of a commodity to a specific mode is a function of the type of commodity and the length of the shipment.

Traffic assignment in a regional transport context is also an area where little research has been undertaken. How do we assign commodity traffic to a rail system when it is the rail carrier's objective to keep traffic on its system as long as possible to enhance its revenue divisions? How should commodity traffic be assigned to the highway system? Do truckers look for only the fastest routes, or is minimizing distance (e.g., fuel consumption) also important? This is an area that has not seen a lot of research in recent years.

Another particularly difficult problem is: How do we evaluate the assignments undertaken? Initially, this study was to examine all traffic flows, personal travel as well as commodity flows. This would have allowed comparison of the assigned traffic with road vehicle counts. Looking at only commodity traffic results in assignments that have no known distribution. Data are available on commercial traffic, but this would include far more types of traffic than the commodities examined here. It is assumed here that the assigned commodities on a highway segment should bear a linear relationship to the total commercial traffic on that highway segment. It is also assumed that such a relationship can be used to estimate total commercial traffic in the future.

The use of statistical analysis to assess the pair-wise interrelationship between observed and modeled highway volumes is very ambitious. The author knows of no transport study at this scale that has used this approach. Urban transport studies of the past were usually satisfied to simply compare the trip length frequency distributions. If the distributions were "similar" the study's modeling was accepted. If the distributions were "dis-similar" usually more data would be collected in the apparent belief that something was missing. No one would think of looking at the actual counts and the assigned (modeled) flows; this was far too rigorous a standard.

In the following section there is an introduction to the commodity traffic generation used here. The reader should be aware that the methods used here are in part from earlier modeling in the Phase 1 report. Nevertheless, all estimates throughout this report are in terms of 1993 commodity tonnages, dollars, and trucks or carloads.

Traffic Generation

As part of this commodity flow study it was necessary to identify the traffic originating and terminating within geographic areas across the United States. When this study began these data were not available for any year after 1977, the year of the last Census of Transportation that included commodity flow data [2]. The more recent data that were available, the Federal Railroad Administration and Interstate Commerce Commission's carload waybill sample, are for the transportation of commodities by rail.

In order to get some idea of what is moving by the various modes, it is necessary to get some idea of the total amount of each commodity that is transported. Two approaches are possible to obtain estimates of total flow by commodity. The first would take the rail traffic available in the ICC waybill sample for each commodity of interest and expand this based on relationships that existed at the time of the 1977 census. For example, if the 1977 census stated that a total of 50,000 tons of commodity x were transported and railroads moved 25,000 tons of this total, this results in an expansion factor of 2 (i.e., 50,000/25,000). One could take the 1993 rail traffic for commodity x based on the ICC sample and expand it by 2 to get the total amount of that commodity moving in 1993.

A second approach is to determine the functional relationships that existed between production and attraction of commodity traffic and key variables capable of statistically explaining these flow variables. It is well established that the total flow of a commodity from a given place is statistically related to the total amount of the commodity produced there. Similarly, total flows to an area are related to measures of local markets. The objective then is to model these productions and attractions.

In order to model productions and attractions, it is necessary once again to work with the flows of 1977. The flows existing at that time are statistically explained by using the levels of related variables at that time. The models derived can then be used with the level of the explanatory variables for 1993 to yield 1993 productions and attractions.

Both of these approaches have recognized shortcomings. In the first approach, there is the assumption that the relationship between the modal share moved by rail maintains a constant relationship to the total amount of the good that is transported. Given the massive rail abandonments in the U.S. since 1977 and deregulation in the motor carrier area, this is a disturbing assumption. In addition to such institutional changes, there are changes occurring in manufacturing processes, e.g., flexible manufacturing and just-in-time delivery practices, that have changed the share of traffic attributable to the different modes. Therefore, it is unlikely that the modal relationships have held constant during the period from 1977 to 1993.

The second approach assumes the relationship between traffic production and industrial production indicators, as well as the relationship between traffic attraction and local market indicators, remain constant over time. If one uses employment as an indicator of industrial productivity, this implies constant productivity per worker for more than fifteen years. Once again the changes in industrial production processes (automation, robotics, computerization) suggest that these linear relationships are probably not stable. One was nevertheless forced to accept one of these approaches since other flow data were unavailable.

This research relied primarily on the second of the two approaches noted above. It utilized multiple regression analysis to develop traffic production and attraction models for each of the nineteen commodity groups examined here.

As noted above, data on the flow of manufactured goods were not available on a current basis when this project started. As a result, this project proposed to estimate these flows based on models derived from the state-level data of 1977. The U.S. Census of Transportation compiled information on the level of tons shipped and received by state manufacturers in that year. These data were compiled and the process of developing traffic generation and production models was undertaken. A few comments are in order on the nature of these models.

The exporting of a manufactured commodity from an area is a function of the level of production of that commodity within the area or its supply. Unfortunately, commodity production data are also not available. Nevertheless, it has been demonstrated repeatedly that an excellent indicator of a sector's production is employment in the sector. Therefore, a key variable in the traffic production models developed is employment in the sector of interest or related sectors. Some of the commodity may never leave the production area since it is consumed locally. To incorporate this tendency, use is made of a population variable to represent this consumer market in several cases.

Flows of manufactured commodities into an area or the attractiveness of an area is a function of the demand for the product. For most manufactured goods there are two markets: the personal consumer market and the industrial market. With regard to the personal market, it is not meant that the manufacturing firms deal directly with consumers; they will most often go through a retailer or wholesaler. Nevertheless, the magnitude of this market is best reflected by the level of local population. The industrial market is often more difficult to identify. As an example, consider a commodity group such as food and kindred products. This group includes all the processed foods consumed by individuals as well as all the ingredients used in preparing other foods. As a result the level of manufacturing in these further stages of manufacturing also represent a market. Once again, employment is used as an indicator of this industrial market.

Returning to the problem at hand, the 1977 production and attraction levels formed the basis for models of the same based on 1977 population estimates derived by the U.S. Census and employment data derived from the 1977 *County Business Patterns* [3]. Models of non-manufactured goods (coal, non-metallic minerals, farm products, and waste) were not developed in the Phase 1 study for reasons previously noted. Models were developed for these sectors here using the 1993 CFS and census data.

The models derived along with an indicator of model accuracy appear as Table 4.1. While other variables important in explaining the levels of production and attraction will no doubt come to mind, there has been a conscious effort made here to keep the variable base limited and readily accessible. All of the models have used only variables on employment by sector, population, or some economic indicators. Forecasting the variables used into the future may be required and all of these have series available from the aforementioned *County Business Patterns*, from population forecasts, or from other government censuses. On a couple of occasions the variables used are a function of other variables estimated. For example, the level of lumber and wood product flows into an area is a function of the level of traffic production in that sector. Derivation of these models yielded a method of estimating traffic produced and attracted by sector for all states of the United States and counties of Indiana in 1993.

Overall the models tend to be accurate based on the adjusted coefficients of determination presented. There are a few of these values that are in the .55 to .70 range; these are low. But the intent here is to get at the major direction and magnitude of the interrelationship. There will always be residuals when one attempts to keep the variable inputs forecastable as well as manageable. It is believed that these models capture the basic relationships reasonably well. It does seem to be worthwhile though to pursue research in the area of commodity traffic generation for future studies.

The appearance of the 1993 commodity flow survey changed the need to use the models derived for estimating state level productions and attractions to some extent. No data were published on the activities at the county level and as a result the models were used to generate Indiana county level productions and attractions.

The 1993 Commodity Flow Survey

As noted previously a commodity flow survey was undertaken in 1993. It was a survey of approximately 200,000 firms in the United States. It was not expected the data would be available for use in this study and this is part of the reason why the alternative methods noted above were developed. As the progress of the study slowed it became apparent that at least some of the data from the survey might be available before the project was over. The United States summary volume appeared in November of 1996 and some state volumes (including Indiana's) have also been published. In January of 1997 a CD-ROM was released by the Bureau of Transport Statistics of the U.S. Department of Transportation that gave among other things data on the amount of commodities produced (in a traffic generation sense) for the nearly all of the

Table 4.1 Models of Production and Attraction

Model Number	Model	Adjusted R^2
(1)	PROD01 = 1445 -.523 AGSER + .0048 CASH	.562
(2)	ATTR01 = .819 PROD01	.660
(3)	PROD11 = 7.6 COAL	.650
(4)	ATTR11 = 3.1 COAL + 5.3 MIN	.657
(5)	PROD14 = .078 MAN	.658
(6)	ATTR14 = .997 PROD14	.977
(7)	PROD20 = .282 FOOD	.940
(8)	ATTR20 = .832 POP + .162 FOOD	.965
(9)	PROD22 = .016 TEX	.931
(10)	ATTR22 = .003 APP + .0001 ALL	.743
(11)	PROD23 = .004 APP	.919
(12)	ATTR23 = .002 APP + .011 POP	.926
(13)	PROD24 = .668 LUM	.808
(14)	ATTR24 = .728 PROD24	.805
(15)	PROD25 = .017 FURN	.906
(16)	ATTR25 = .033 POP + .002 FURN	.960
(17)	PROD26 = .103 PULP + .056 LUM	.886
(18)	ATTR26 = .085 PULP + .259 POP	.953
(19)	PROD28 = .150 CHEM + 1.164 PET	.758
(20)	ATTR28 = .077 CHEM + .455 PET +.683 POP	.851
(21)	PROD29 = 6.857 PET	.945
(22)	ATTR29 = 4.007 PET + 1.881 POP	.938

(23)	PROD32 = 2.882 POP	.851
(24)	ATTR32 = 2.914 POP	.871
(25)	PROD33 = .085 MET	.982
(26)	ATTR33 = .093 MET + .061 FAB	.923
(27)	PROD34 = .013 MET + .034 FAB	.927
(28)	ATTR34 = .035 FAB	.861
(29)	PROD35 = .013 MAC	.883
(30)	ATTR35 = .010 MAC	.878
(31)	PROD36 = .004 MET + .004 FAB + .003 ELEC	.826
(32)	ATTR36 = .005 FAB + .034 POP	.915
(33)	PROD37 = .040 TRAN	.753
(34)	ATTR37 = .027 TRAN	.837
(35)	PROD40 = .00048 POP	.704
(36)	ATTR40 = .0067 MAN	.791
(37)	PROD50 = 1.097 ATTR50	.858
(38)	ATTR50 = .245 POP	.857

Notes: Most of the explanatory variables above are employment in specific STCC (SIC) industrial classes according to County Business Patterns. Exceptions are the PROD and ATTR variables which represent tons of product shipped or received by STCC, e.g., PROD01 is the tons of farm products shipped and ATTR01 is the tons of farm products received. Other variables are defined as follows: AGSER = employment in SIC 07; ALL = total employment; APP = employment in SIC 23; CASH = gross cash receipts (in $1,000s) from farming; CHEM = employment in SIC 28; COAL = employment in SIC 11; ELEC = employment in SIC 36; FAB = employment in SIC 34; FOOD = employment in SIC 20; FURN = employment in SIC 25; LUM = employment in SIC 24; MAC = employment in SIC 35; MAN = total employment in Manufacturing, SIC 2 and SIC 3; MET = employment in SIC 33; MIN = employment in SIC 14; PET = employment in SIC 29; POP = total population; PULP = employment in SIC 26; TEX = employment in SIC 22; TRAN = employment in SIC 37.

industrial sectors of interest here.

After considering the quality of the data being released it was clear that the 1993 data being released was of a much higher quality than the 1977 Census of Transportation data. This was due in large part to the fact that the 1977 commodity flow data was based on a sample of 20,000 shippers and the 1993 data was based on a survey of 200,000 firms. The latter study is also much more aware of the statistical nature of the data collected, e.g., coefficients of variation are presented for most data and "unstable" data (usually based on small numbers in the sample) are not published. It seemed logical to use production and attraction data from the 1993 flows if this was at all possible.

A portion of the data released to date makes it very clear that flows are examined primarily from the traffic origination side. Tables are available that give traffic production by STCC codes for the industrial production of interest here. These data were extracted and used in their published form.

Traffic attractions presented more of a problem. Recall that all of the state volumes have not been released. It is not clear that this would solve the problem or not. The problem quite simply is that there do not appear to be any figures given for total attractions by state and industry. There are tables on the CD-ROM that yield flows from an origin state to destination states by commodity and this might appear to yield a route to the data of interest, i.e., one could add the flows of each commodity of interest from all states to the destination state and get total attractions. This would be quite possible if all the data appeared on the CD-ROM, but interstate flows are often very scarce and the data are withheld for proprietary reasons. These same proprietary concerns would not enter into consideration if the data included the total traffic attracted by industrial sector. In other words the Bureau of Transport Statistics has the data and could do a special aggregation of the data for state modeling purposes.

There was not sufficient time to pursue negotiations for the release of attraction data by the Bureau of Transport Statistics. Instead two pieces of aggregate information were used along with the regression models previously noted to estimate the traffic attracted to destinations. In the first case the models were run to estimate the "volume" that would be attracted to each state. The sum of these estimates by commodity was equated with the total attractions by commodity for the country; the latter total attractions is one of the pieces of aggregate information that kept the system in line with actual data.

The second piece of information was the total traffic attracted by commodity for the state of Indiana. This piece of data was used in the same manner as the other national data were used. In effect, the aforementioned models were run to estimate attractions for Indiana counties and the total attraction for the state became the flow limit for allocating commodity traffic to destination areas.

The result of these various operations can be summarized succinctly as follows: the total flows produced by the states are equal to the total flow produced by the nation; the total flows attracted by the states are equal to the total flow attracted by the nation; these same statements also apply to the counties of Indiana and the sum of their productions and attractions are equal to these values for the state. These controls enhance the accuracy of the methods used here.

Traffic Distribution

The distribution of traffic in the Phase 1 report was accomplished by developing several statistical models for traffic generation, as described in the previous section, and using these models to identify the production and attraction vectors for input to a fully constrained gravity model. The latter model is also sometimes called an entropy model (see Wilson [4] and [5]). Such a model can, given productions and attractions and an average length of shipment, yield estimates of the flow between all origins and destinations. It does this in such a way that the estimated traffic volume from any origin is equal to the value inputted to the model, the estimated traffic volume to any destination is equal to the value inputted to the model, and, in general the average shipping distance of the estimated flows is equal to the value inputted to the model. In effect, the outcome is constrained to meet all of the initial input parameters of the problem. This tends to yield the most accurate results of any general flow model currently in use.[1] One could alter the output of any model and get more accurate results. But the basis for such alterations is not clear, nor is it usually defensible.

This model has the general form:

$$S_{jk} = A_j B_k O_j D_k \exp(-\beta c_{jk})$$

where S_{jk} = the amount of a given commodity shipped from origin area j to destination area k;

O_j = the amount of a given commodity available for shipment at origin j;

D_k = the amount of a given commodity demanded by destination k;

c_{jk} = a measure of the cost or impedance of moving from j to k.

[1] Some research was undertaken at the beginning of this project to determine whether artificial neural network models would yield results that were more accurate than the fully constrained gravity model used here. Initial results working with small flow matrices were very promising (see [6]). There does seem to be some difficulty in fitting these models to large sets of data of the type examined here. This statement is based primarily on unpublished research undertaken by Garth Banninga on solid waste flows in Indiana and by Jirong Xie on passenger travel on the Amtrak system.

In addition,

$$A_j = [\Sigma\ B_k\ D_k\ exp\ (-\beta\ c_{j\,k})]^{-1}$$

and

$$B_k = [\Sigma\ A_j\ O_j\ exp\ (-\beta\ c_{j\,k})]^{-1}$$

The above formulation is rather straightforward. Some comments are in order on the impedance or cost of movement factor, c_{jk}. This factor is defined here as the distance between a location *j* and a location *k*. Obviously other functions could be used, but this seemed as good as any [7]. The purpose of the impedance factor is to exercise a negative influence on interaction at increasing distances, other things being equal. Some researchers believe that this function should be actual road distance between places. If it were this would certainly complicate subsequent analysis which in part tries to identify the need for additional roads.

The distribution model above is available in the form of a FORTRAN computer program that operates on desktop personal computers. The FORTRAN code for this program appears in the appendix as GUNNAR5; it was prepared during Phase 1 of this project.

A few comments are in order about accuracy of this fully constrained gravity model. The model should not be viewed as something that is capable of a perfect replication of a set of existing flows or shipments. To begin with, we rarely know what the actual flows look like. Actual flows are almost always based on a sample of cases that is expanded, e.g., the rail carload waybill sample (a proportional sample of 1% to 4%), the 1993 commodity flow survey (a sample of the shipping practices of 200,000 firms), airline traffic city pair data (10% of all tickets sold), commuting flows of the 1990 census (1 in 6 households), and others. As a result even the "actual" data is subject to considerable error and it would be unreasonable to expect this data to be perfectly reflected by the model.

Instead, the model should be viewed as capable of replicating major shipping patterns in the area of interest; this is the United States and Indiana in the present instance. In effect, if certain aspects of the flow are known this model should yield what could be called the most probable set of flows given this information. This is sometimes called the most probable macrostate. This yields the major patterns of flow, which should be the major concern of the analysis.

As noted above several delays in the project slowed its progress to the point that reports began to appear for the 1993 Commodity Flow Survey. When the project began it was expected that it would be completed long before these data were available. This was part of the reason for using the modeling approach to estimate productions and attractions at the state level for Phase 2. Given that the primary concern here were the flows of 1993, it seemed reasonable that these 1993 Commodity Flow Survey results should be included as much as possible.

As a result the study took the state-level traffic produced by each manufacturing sector of interest here. These data have not all been published at this time, but they are available by state of origin (for the flows) on a CD-ROM released by the Bureau of Transportation Statistics in 1997. Terminating traffic as previously noted is not available for individual states in any form at this time. It was assumed that the sum of the flows originating by commodity in the states of the U.S. is equal to the sum of the flows terminating by commodity in those states. Let's examine this assumption a little further.

The 1993 Commodity Flow Survey examined shipping data supplied by 200,000 manufacturing firms in the United States. These firms supplied information on their shipping to all possible destinations in the U.S. and the destinations included consumer markets and industrial markets. If goods were being shipped to California for possible export via Long Beach or Los Angeles, they appear in the data as shipments to California. On the other hand there are some flows that originate outside the U.S. and are imported for sale in the U.S. market. Since these latter shippers are outside the country they are excluded from the survey. The flow survey is moot on the potential volume of flow that this might include.

In 1993 the United States exported \$465 billion in goods and it imported \$581 billion in goods. However, for manufactured goods the situation is worse than this. As of 1995 U.S. exports were valued at \$452 billion while imports were at \$630 billion. By assuming that total traffic terminating is equal to total traffic originating we will pick up \$452 billion dollars in imports that otherwise would be lost to the analysis. There will still be \$178 billion dollars in traffic that is excluded from the analysis. In addition, it is not possible to assume that the goods being imported are equal to the goods being exported. In 1995 the U.S. exported \$15 billion in automobiles, while it imported \$50 billion worth of automobiles. Nevertheless, the assumption of equality picks up a substantial amount of traffic that would otherwise be lost.

The project also used actual data for Indiana to refine the modeled estimates for the counties of the state, i.e., the sum of the county estimates for each category was set equal to the state total. As a result the following statements can be made with regard to commodity flows examined in this study:

> (1) Total flows from all states as used by the gravity model are equal to actual traffic productions by manufacturing category for

those states.
(2) Total flows from Indiana and total flows to Indiana, by commodity, as generated by the model are equal to the actual flows as given in the commodity census.
(3) The sum of the total flows as generated by the states for productions and attractions are equal to national totals for these.

One might reasonably ask what can differ between the actual and the estimated traffic. One thing is the modal split in traffic. Although the earlier Phase 1 study used several modal categories, this Phase 2 study uses an extensive array of modal categories and assigns traffic to modes based on what it is and the length of move involved. These assignments are based on traffic by distance categories as published in the 1993 census. This could account for some minor errors in the analysis.

A second possible source of error may be due to the values used for the average length of shipments as used in the gravity model. Recall that this is one of the constraints used by the fully-constrained gravity model in fitting the flow data. The 1993 commodity census does publish an average shipment distance for each commodity. However, the problems with this should be apparent. Let us assume that a shirt manufacturer ships a gross of shirts to a store in Indianapolis. The average length of the shipment is the distance from the manufacturer's plant to Indianapolis. Let us assume this distance is 800 miles. Now let us assume that an individual in the Indiana capitol orders a shirt from a different manufacturer that is 400 miles away. The average length of the shipment is 800 miles plus 400 miles divided by two shipments, or 600 miles. This is obviously not what we want.

This method of calculating the average shipment distance created significant problems for the census in the manufacturing category of printed matter. It resulted in all copies of a weekly news magazine being shipped to a state distributor for retail sales counting as one shipment, while an individual's mailed copy would also be counted as a single shipment. The census did not release any data on state activity in this category.

In order to overcome this problem a decision was made to examine average shipping distance per ton of a given commodity. The census publishes data on the total amount shipped in tons and the total ton-miles for each commodity. Dividing total ton-miles by total tons gives an average shipping distance per ton of the commodity. This seemed like a perfectly reasonable value and so it was the average shipping distance used in the gravity model. But there were still two additional problems.

The first of these is a problem of geographic scale. The total tons and ton-miles are based on actual distances moved. This planning effort abstracts from that reality and says that for most states there is a single origin and destination for commodity shipments in the state. This can create a problem. To illustrate it, assume only four western states trade a commodity and the distance between their centroids (the single origin and destination points) ranges from 400 to 500 miles. If all of the actual parties making and selling this product are located within 200 or 300 miles of each other, then the average shipping distance per ton will be somewhere between 200 and 300. There is no way that a gravity model could be fitted to these data since all possible interstate shipment distances would be too great. The model might assume that each of the four states traded only with itself resulting in an average shipment distance of 200 to 300 miles. It would be possible to calibrate this model, but this would not be an accurate representation of the flow occurring.

A second problem was that there was no way to be certain that a gravity model calibrated to state and Indiana county data would represent Indiana flows in an adequate way. The flows generated are a type of average picture of what is happening in the country. In a similar manner we can say the per capita income for the United States is a reasonable average, which when expanded by total population yields total personal income for the nation. However, it may not give a reasonable indication of what the per capita incomes are in the counties of Mississippi. So although we may get a reasonable picture for the nation it may not give us a reasonable picture of flows in the state. The question was how to overcome this problem or put another way, how could we be reasonably sure the flows were approximately what is going on for Indiana.

The solution to this problem was found in the 1993 commodity census. That census gives very good data on Indiana as an origin of shipments for the commodities examined here. It gives tons and ton-miles and as a result it is possible to look at the average length of shipments from Indiana by commodity ton. In effect, the gravity model was fitted to the national system of flows and the average shipping distance per ton was calculated as part of this process. The program used was modified to then calculate the average shipping distance per ton for all goods from Indiana to all of its destinations. This in turn was checked against the average shipping distance per ton of commodity for Indiana. Equality of the two means was not expected, but significant differences would lead to some adjustment to the national figure in an attempt to bring the means closer together.

The actual average shipping distance per ton for the nation and Indiana and the model generated averages are given in Table 4.2. As the reader can see the numbers are reasonably close. One might be tempted to try to perfectly replicate the national figures, but the reader should bear in mind that the commodity census is a sample of 200,000 firms, and it could also be subject to error in data reported by the firms.

The primary item being modeled here is tonnage of commodities shipped between origins and destinations in the United States for the year of 1993. As a result these flows are annual tonnages. In order to have something to compare average daily traffic with, it was necessary to reduce the flows to a daily basis. An examination of the *Highway Capacity Manual* [8] revealed

Table 4.2 Actual and Modeled Average Shipment Distance per Ton of Commodity

Commodity STCC	Actual U.S. Average	Modeled U.S. Average	Actual Indiana Average	Modeled Indiana Average
01	434	434	435	432
11	432	432	85	436
14	87	116	44	122
20	315	311	333	311
22	458	445	236	489
23	658	420	391	397
24	182	190	220	222
25	591	592	794	563
26	464	313	313	314
28	434	345	280	294
29	152	153	89	140
32	105	202	124	189
33	365	365	356	361
34	359	358	342	345
35	559	500	472	473
36	649	505	481	483
37	560	487	449	446
40	211	211	181	243
50	560	507	426	465

Notes: The actual average distances were derived from ton-mile and tonnage data in the 1993 CFS. They differ from the published average distances which are average length of shipments. Shipments may be considerably less than a ton in many cases for manufactured products. The Indiana average is the length of "originated" shipments from Indiana derived using ton-miles and tonnages as above.

that truck traffic generation was approximately equal from Monday through Friday, but on the week end the truck traffic generation rate appears to be about 44%. Adding five days and two days at .44 each yields 5.88 days per week or 306 days per year of trucking. Therefore, dividing the total annual flow by 306 yields a good estimate of commodity truck trips per weekday. Multiplying this weekday rate by .44 gives a reasonable estimate of the traffic on a Saturday or Sunday.

Modal Split

Once traffic is distributed between origins and destinations there remains a question of the modes selected for the movement of that traffic. In the Phase 1 report this modal split was handled by examining the modes used for the movement of product-specific traffic at the time of the 1977 Census of Transportation. These modal proportions were then used to assign traffic to specific modes. The available modes were limited in the 1977 census and consisted of rail, motor carrier, air freight, parcel, water transport, and pipelines. When the Phase 2 project started the project staff was going to use the same computer program and the same categories for allocating the 1993 flows. The project staff had concerns about this, but no reasonable alternative was available.

The concerns were primarily that the transport environment had changed significantly since 1977. Notably, the motor carrier and railroad industries had been deregulated and this has created numerous changes in these modes. For example, there are far more independent carriers in motor carrier transport today than there were in 1977, and backhauling is practiced far more than it once was. It was assumed that several of these changes could be incorporated in the analysis by minor changes in the modal assignment weights, but such changes would lack rigorous definition and justification, and were viewed as undesirable.

Delay of the project by network definition problems resulted in the initial publication of the 1993 Commodity Flow Census volumes during the last several months of the project. Among those volumes was the United States summary; this includes the proportion of traffic shipped by various modal combinations for different distances. A decision was made to update the modal split model and incorporate new weights from the 1993 census.

The new modal split computer model was named NEWMODE and a listing of it appears in Appendix C of this report. Rather than a half dozen modes this 1993 data includes data on nine single modes and eight multiple mode categories as indicated by Table 4.3. While additional detail is always desirable it sometimes creates problems as well. The primary areas of interest in this document are highway and rail traffic. It would simplify the project if only truck and rail were given, but this would ignore several modal combinations that seemed to be rather common, e.g., truck and rail or truck and air. So that this traffic would not be lost all seventeen categories

Table 4.3 Modal Categories for Traffic Split

Single Modes
Parcel, U.S. Postal Service, or courier
Private truck
For-hire truck
Air
Rail
Inland water
Great Lakes
Deep sea water
Pipeline
Multiple Modes
Private truck and for-hire truck
Truck and air
Truck and rail
Truck and water
Truck and pipeline
Rail and water
Inland water and Great Lakes
Inland water and deep sea

were used in the modal split. In addition, the two principal categories of highway and rail also included the other modes such as truck and rail in recognition that sometimes the trucking may take place on the origin side of the trip and sometimes it may take place on the destination side of the commodity move. This created an obvious problem of over-counting since the traffic was attributed to both rail and highway. This overlap was minor, but it is the reason for a slight discrepancy between the sum of the traffic and the sum of the traffic by modal categories.

The computer program NEWMODE splits traffic by examining the lengths of the shipment and knowing the product of interest. In other words each of the fifteen commodity groups examined here has a set of distances (less than 50 miles, 50 to 99 miles, 100 to 249 miles, 250 to 499 miles, 500 to 749 miles, 750 to 999 miles, 1000 to 1499, 1500 to 1999, and 2000 or more miles) and for each distance group there are seventeen modal categories, plus some summary classes such as highway, or rail. Let us assume we are looking at 1,000 tons of primary metal products being shipped 800 miles. The shipment has a .626 probability of moving by highway, a .366 probability of moving by rail, and .008 probability of moving by an unknown mode (due to errors in reporting). NEWMODE assigns 626 tons to a highway mode, and 366 tons to a rail mode. The remaining 8 tons are ignored since they can not be assigned to a mode with any confidence. Obviously, if this shipment is from a single plant the move will go by either rail or highway, and it is unlikely that it would be split in this fashion. However, over the millions of tons shipped this allocation procedure would be capable of replicating the flows that did occur in 1993 based on the reported census data.

Commodity Density

The modal split to this point has dealt exclusively with tons of commodities. However, our major interest is in motor carriers and rail cars. In other words it is necessary to assign the traffic to vehicles representing each of these modes. Unfortunately, we must move away from the 1993 census data since it is moot on the question of how many tons of different commodities will fit into a rail car or a tractor trailer motor carrier.

The volume of a commodity that will fit into a given space is its commodity density. Density values for the Phase 1 report were obtained from a 1976 Interstate Commerce Commission study. However, that source was also viewed as unreliable since rail cars have increased in size, as have motor carriers. A more recent source of data was seen as the 1993 carload waybill sample.

To obtain new density factors the tonnages of the 19 commodities coming into Indiana, and leaving Indiana by rail according to the expanded Waybill Sample were aggregated by commodity and rail carloads. Division of the former by the latter yields tons, by commodity, per carload, or commodity density. As one might expect these density factors differ based on whether they are in inbound or outbound and this may reflect the commodity's stage in the manufacturing process. A weighted average of these inbound and outbound density factors was calculated for rail cars and these appear as Table 4.4.

The density factor was estimated for STCC 23 (Apparel) since none of that commodity appears to have moved by rail. In addition the factor for STCC 25 (Furniture and Fixtures) is based only on the export (outbound) traffic since there was no inbound traffic for this sector.

The density factors for motor carrier traffic assume that a rail car can handle 100 tons and a truck can carry 40 tons or 80,000 pounds, or 40% of the same product. In other words it was assumed that the motor carriers could handle 40% of the density factor of a rail car by commodity, and this is the source of the motor carrier density factor in the table. One could argue that some states permit vehicles with weights in excess of 80,000 pounds, but there is also a considerable amount of highway traffic moving in vehicles smaller than this. As a result, assuming all motor carriers are 40 ton vehicles seems a reasonable standard for use here.

Two other computer programs were written to handle the allocation of tonnages by rail and highway motor carriers. These programs, entitled ALLORWY and ALLOHWY (see Appendix C), read data from the output of NEWMODE and create a set of flows by origin and destination and mode. These files can be used to create a flow matrix which can then be assigned to the appropriate transport network.

In addition to translating the tons into rail cars or motor carriers, ALLORWY and ALLOHWY also create files of tons and dollars. The former could be used as a point of departure for other traffic analyses of interest to the state. The latter after it is assigned to the network enables planners to know the dollars by commodity coming into the state or leaving the state, by route and mode. These monetary values are derived from average values per ton of commodity according to the 1993 Commodity Flow Census (see Table 4.5). These may also be useful for further analyses.

Mail Density

It was noted in an earlier chapter that estimates have been made of the amount of personal mail and express mail moving between Indiana counties, as well as between those counties and the rest of the United States. It is necessary to know the density of mail in order to determine the number of vehicles that would be involved in its transport. This is only part of the problem, but let us address it first.

Mail arriving in your local community usually arrives by a contact mail carrier operating a tractor-trailer or semi. Included in the trailer are bags of mail that can weigh no more than 70 pounds. Most are filled to this limit. A typical trailer can hold an estimated 450 bags or 31,500 pounds of mail. This translates into 15.75 tons per motor carrier. This is a maximum rather than

Table 4.4 Traffic Density Factors for Rail Cars and Motor Carriers by Commodity

Commodity STCC	Import rail traffic	Export rail traffic	Weighted rail density (tons)	Weighted truck density (tons)
01	94.90	96.20	96.13	38.44
11	100.60	99.10	100.42	40.17
14	97.10	97.40	97.20	38.88
20	77.35	80.36	79.52	31.81
22	25.00	15.00	18.33	7.33
23	-------	-------	*10.00	*4.00
24	73.88	55.50	72.27	28.91
25	-------	15.00	15.00	6.00
26	64.82	50.64	62.10	24.84
28	85.11	90.11	87.58	35.03
29	63.20	77.16	65.90	26.36
32	86.70	77.10	81.15	32.46
33	87.48	85.21	85.82	34.33
34	28.40	16.16	19.76	7.90
35	68.75	21.70	28.42	11.37
36	18.80	16.25	16.69	6.68
37	19.93	23.40	22.50	9.00
40	75.40	82.60	78.47	31.39
**50	92.85	14.88	86.56	34.62

* Estimated values

** There is no STCC 50. It is used here to represent STCC 21, 27, 30, 31, 38 and 39.

Table 4.5 Commodity Value per Ton

Commodity STCC	Value per Ton
01	$ 224
11	21
14	12
20	997
22	4128
23	19252
24	191
25	4193
26	898
28	977
29	191
32	114
33	858
34	2795
35	12954
36	13630
37	7447
40	139
50	7855

a minimum because the mail comes to your local community in theory if there are only a few pieces of mail to be delivered there. So all communities get one motor carrier as a minimum and the maximum is determined by the number of motor carriers necessary to move the mail with a limit of 15.75 tons each. It is assumed that the same operating scenario also applies to private express mail companies.

Some counties may receive considerably more trucks than estimated here. That does not mean that our estimates are off. We are looking only at non-commercial mail (letters, manuscripts, photographs from your children or parents, contracts and the like). Other mail that may include commercial materials, e.g., a shirt from a mail order house, or some fresh fruit from a popular West Coast mail order firm, are included as parcel moves in the commodity flows examined elsewhere in this report.

Traffic Assignment

This section describes the procedures used in the traffic assignment portion of the study. It should be noted at the outset that the purpose of traffic assignment is to *assign* flows that exist or flows that have been predicted or forecasted by a model to the transportation network of interest. In the present case the networks of interest are primarily the highway and rail networks of the United States with particular interest in these networks within Indiana.

Methods of assigning traffic are numerous. The simplest of these methods is referred to as "all or nothing" assignment. In this procedure the methodology assigns traffic moving between some area j and some area k to the shortest path (route) between this origin and destination. All possible pairs of origins and destinations have their traffic assigned in exactly the same manner. There is no consideration given to the capacity of links in the paths selected or whether travel time on the links will be affected by congestion.

"All or nothing" traffic assignment has more than a few critics against its use in urban transportation planning, however the focus here is on regional or statewide transportation planning and most of the criticisms seem inappropriate in the regional context. For example, nearly everyone in Indianapolis that found they suddenly had to drive to Chicago would take Interstate 65 simply because of the shortest travel time of this route. In an urban context there might be some question about your route in getting to this interstate highway, and probably other assignment procedures would be appropriate for this portion of the trip. It does not matter in this study because the flows being examined are intercounty and interstate trips.

Another traffic assignment technique is called "capacity restraint" assignment. In this case one can make use of a capacity limit on traffic to be assigned to the links of a path between an origin and destination. The assignment procedure uses the same "shortest path" approach that the "all or nothing" approach uses, but once a link's capacity is reached traffic is diverted to the "second shortest path," and then on to the "third shortest path" having the capacity to handle the traffic. The recognition of capacity as having an impact on route selection is a positive attribute of this method.

A third traffic assignment approach is "stochastic user equilibrium" or SUE. This approach is not deterministic like the previous two methods, but rather allows users of a transport network to vary their behavior. All users are faced with decisions as to what routes to take between an origin and destination of interest. Each of these routes has a certain probability of being selected based on capacity, travel time, congestion, or whatever variables are specified by the modeler. Selecting one of these routes may lead to congestion and increases in travel time, so the situation faced by the next traffic to be assigned is different from the situation encountered by the previously assigned traffic. Theorists find this approach to traffic assignment more attractive because it recognized the choice nature of urban travel. In a regional context where congestion and capacity are rarely problems the method has less appeal.

There are numerous variations on these traffic assignment techniques, but these are representative of the approaches in use. Each of these was evaluated for use in the present study. In addition there are several steps common to each of these techniques. Let us briefly examine the process further and the results of this effort.

The Highway Network and Cost of Movement

Each of the traffic assignment techniques requires the construction of a network over which movement can take place. This network connects all origins to all destinations and includes the "cost" of movement over the links and in some cases the capacity of the links to hold traffic. Cost may be a misleading term because the measure used is rarely in dollars and cents. Instead studies over the years have used distance, travel time, or traffic flow functions related to distance or travel time. This project used travel time as its initial measure of travel cost. For large scale studies over an area the size of the United States travel time is rarely known. Instead it is approximated by the following:

$$Traveltime = (Length) / (Speed)$$

Here the length is in miles and the speed is in miles per hour. This results in travel time being measured in hours or parts of hours.

Until quite recently the highest speed throughout the United States was found on the Interstate Highway System and it was 65 miles per hour. This was the case in 1993, the year of this study. Unfortunately from a modeling point of view, the next speed found in most states was

55 miles per hour. This had the effect of making 65 mile per hour highways very attractive in most traffic assignment procedures. Put another way, models that seek to lean toward shortest path solutions in assignment of traffic find the Interstate Highway System to include these paths and this results in the bulk of the traffic being assigned to these links. Given the network defined for this study - a detailed state network connected to a circular regional network, and an Interstate Highway network connecting these to the rest of the United States - this may have resulted in some bias, i.e., flows from outside the circular region were assigned to the Interstate Highway System and this is the way they entered that region.

One final point on the network is that whenever the cost of travel, or the way in which it is being measured, changes, or whenever the links in the network change for some reason, it is necessary to generate a new network for assignment purposes since the network expects to move flows between centroids, the network nodes must be consistent with the defined network.

Target Flows

The flows used by the traffic assignment procedures were the flows for all goods examined in this study. In other words, it included shipping the total highway tonnage of all 15 manufacturing groups and the four resource based commodities and two types of mail included in this study. This represented the sum of the 21 gravity model distributions by O-D pair. The reason for evaluating the traffic assignment routine using total flow was that this was the only variable that came close to existing data on actual flows. Existing traffic count data are actually expanded numbers of commercial vehicles per day on Indiana's highways over the period from 1991 to 1994. This became the target to which the traffic assignment had to demonstrate a relationship.

It goes without saying that commercial vehicle count data is not the best measure of manufactured or primary commodity traffic on the highways. After all commercial traffic includes the movement of delivery and large service vehicles, as well as empty trucks. We have not examined the empty tractor trailer combinations here. The major implication of this is that the target flows were larger than the flows to be assigned. In addition, this study looked at intercounty flows, not intracounty flows, and as a result the target flows were generally larger within an origin or destination county.

One other point noted above is that the digital highway network of Indiana used here is not as complicated as the actual highway network. Flows generated by models must be assigned to the digital network and this may result in higher flow volumes than the actual flows observed on highways of the state, since the digital network excludes thousands of miles of lesser roads that in reality may move some traffic.

The above points are not open to argument or debate; they are simply reasons why we generated in this assignment appear as Figures 4.1 and 4.2. The first map shows the traffic volume as reflected by the width of the bands on the highways. The second map shows these volumes numerically. Figure 4.3 illustrates the band width appearance of the national flows. This latter map would not be accurate outside the Indiana region due to the minimal number of external nodes, however it does illustrate the way these flows reach the state of Indiana.

Evaluation of traffic assignment results is not an easy matter. A researcher's first impulse is to simply undertake a correlation and regression analysis of the statistical relationship between the assigned flows and the target flows from the total commercial vehicle road counts. In general this is not done in transportation planning studies at the scale of this study. Instead planners look at the distribution of trips and retain the assignment if it is "close" to the observed distribution. Nevertheless, a statistical analysis was undertaken here with that analysis demonstrating no significant relationship between these two variables.

This was surprising initially, but an examination of the patterns of flows from the actual road counts clarified the picture somewhat. As noted the actual flows (road counts) include far more than the manufactured goods examined here. The latter do not include the delivery of goods or products to distribution or retail centers. This results in significantly higher volumes around urban areas for urban goods delivery. In addition, there are several types of commercial traffic that are not included here. An examination of the road counts data collection sites also revealed a large number of these located within urban areas, which would result in less of a relationship with the assigned flows.

The original research design to be used here called for the comparison of assigned flows with the road counts at locations around the states boundary. The more complex network, the 1993 Commodity Flow Survey, and the availability of a much larger and more detailed database of road counts led to the modification of that design. In addition, the original design would have included "all" travel which would have yielded more stable flows. It was expected that inclusion of all flows would have resulted in assigned flows that would be larger than the expected (actual) flows based on road counts, since the network to which the traffic would be assigned was much less complete than the final network adopted for use here.

It still seemed desirable to do some type of statistical comparison of the assigned flows and the road counts data. A sample of 40 locations in rural areas of the state, but including all types of highways, was drawn. A map of the location of the data collection sites for these data appears as Figure 4.4. A table of the relevant data for these sites appears as Table 4.6. There is some clustering of the sites, but the sample counts in these cases were for different roads or highways in these cases.

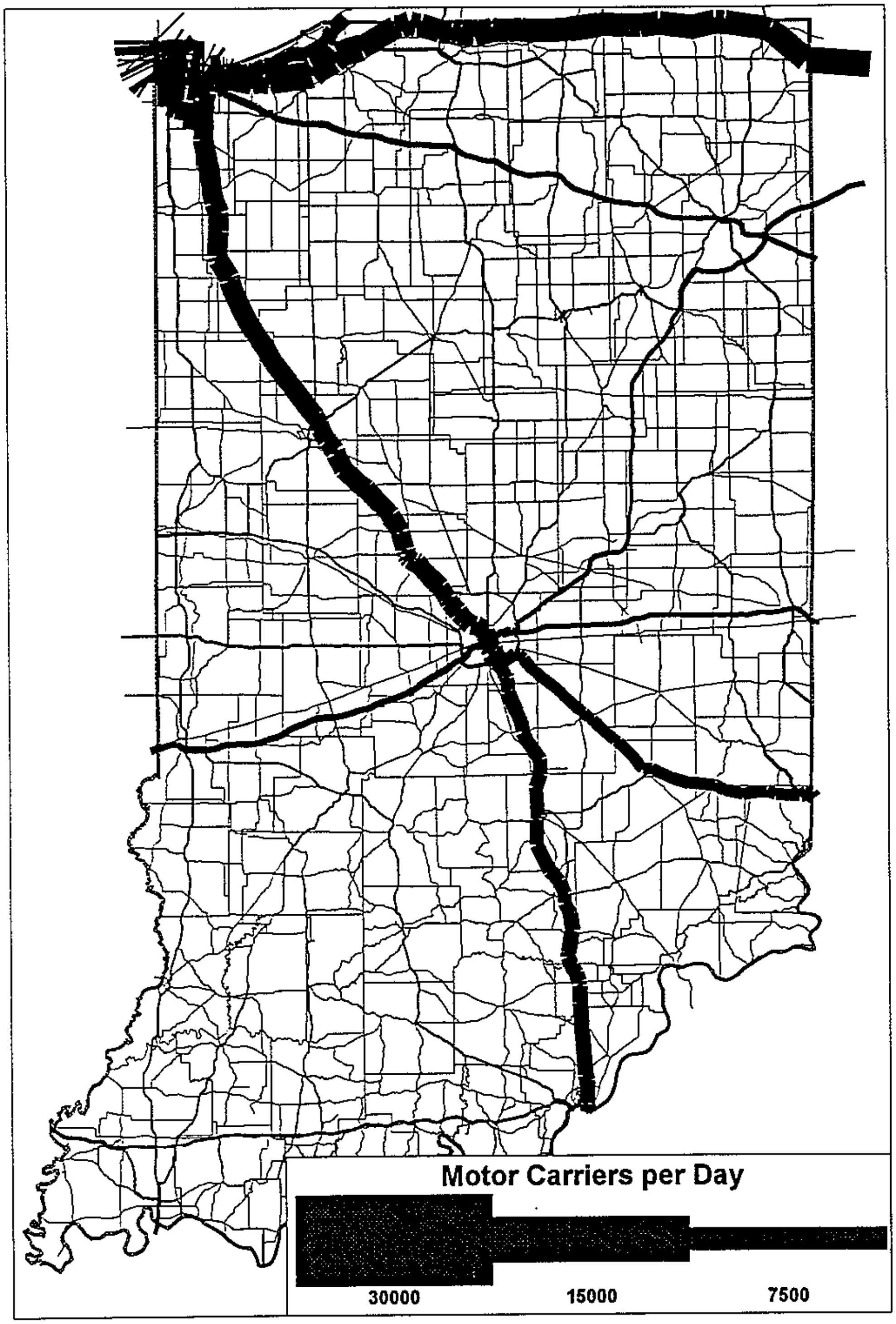

Figure 4.1 The Assignment of Total Daily Truck Traffic to Indiana Highways

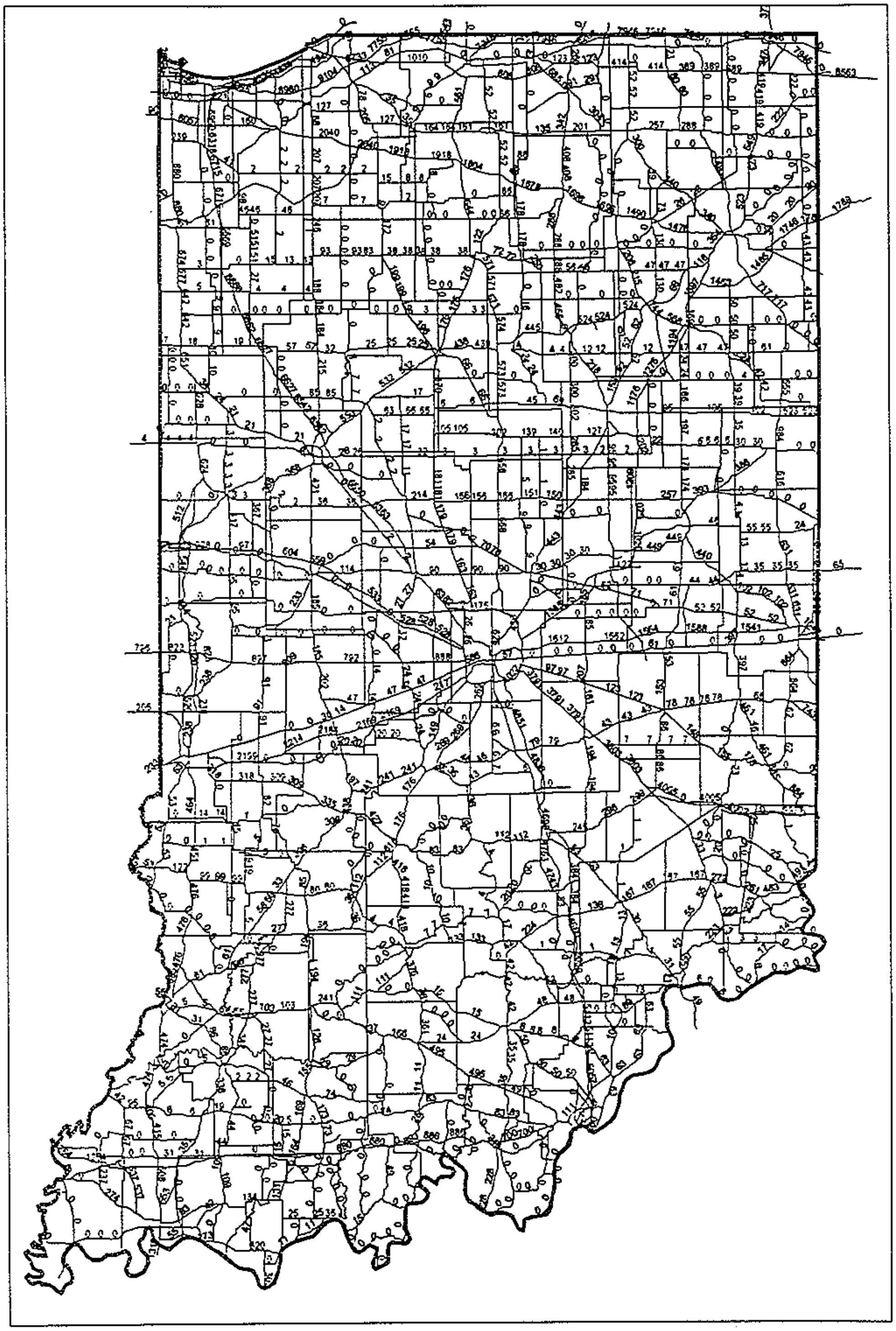

Figure 4.2 Daily Traffic Volumes Based on Modeling

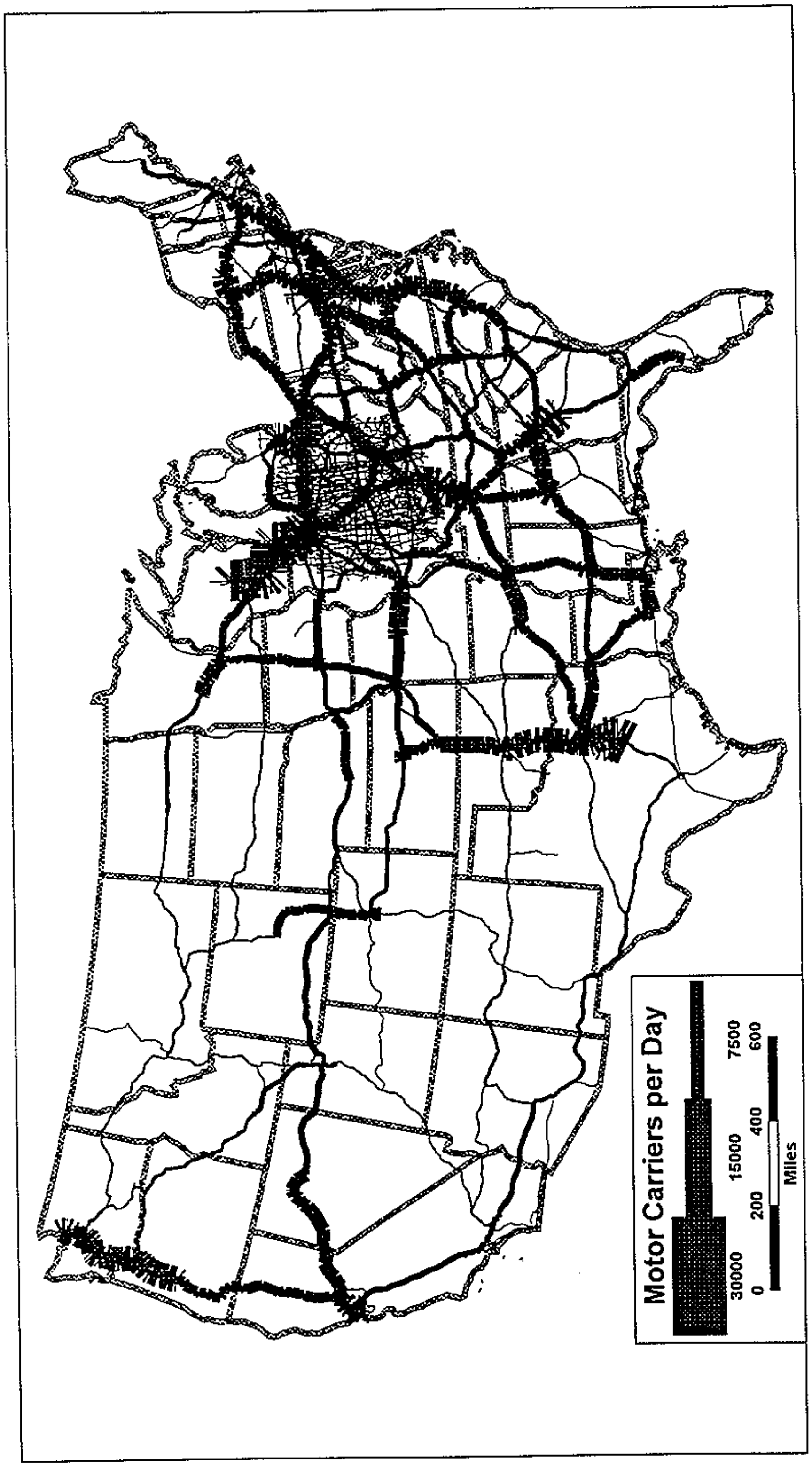

Figure 4.3 An Assignment of the Total Daily Truck Traffic to the National Highway Network

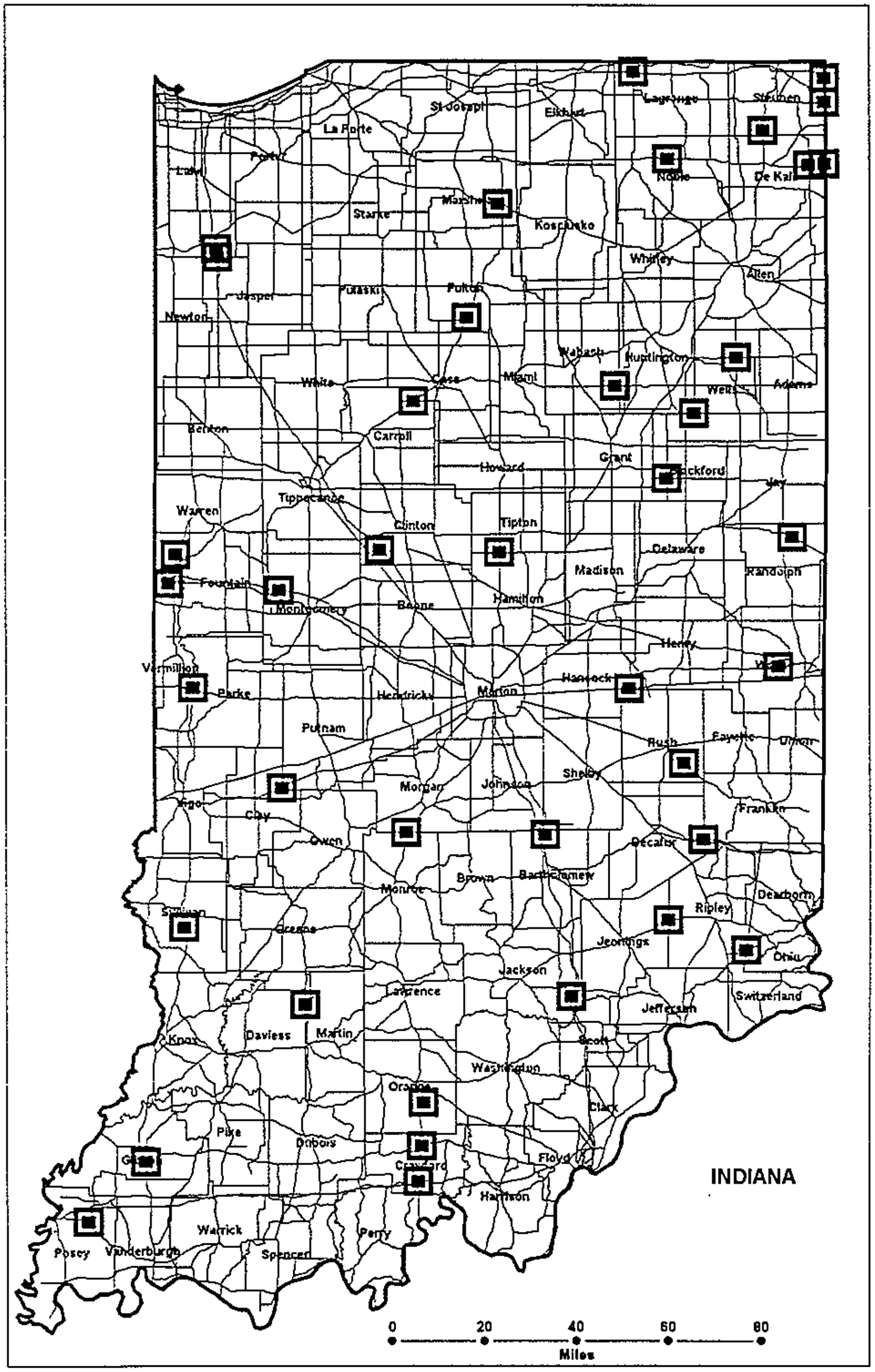

Figure 4.4 Data Sites Used for Evaluating the Modeled Daily Motor Carrier Traffic

Table 4.6 Highway Locations of the 40 Data Sites

Station ID	Location	Highway
4401	ON SR 66 0.50 MI W OF SR 165	ST-066-65-01
4737	ON US 41 4.00 MI S OF BR 4616/SR 64	US-041-26-01
6201	ON US 41 0090 MI S OF OLD SR 54 LT	US-041-77-01
7057	ON US 36 0.11 MI E OF VERMILLION CO LINE	US-036-61-01
8137	ON I-64 6.43 MI E OF DUBOIS C/L	INTST-064-13-01
8313	ON SR 64 1.44 MI E OF SR 37	ST-064-13-01
8337	ON SR 37 7.31 MI N OF CRAWFORD C/L	ST-037-59-01
8737	ON US 231 0.50 MI S OF SR 645	US-231-51-01
9337	ON I-70 0.10 MI E OF CLAY CO LINE	INTST-070-67-01
9633	ON SR 37 0.10 MI NORTH OF MONROE CO LINE	ST-037-55-01
10993	ON I-74 1.00 MI W OF SR 63	INTST-074-83-01
11057	ON SR 63 4 MI N OF SR 263	ST-063-86-01
12521	ON SR 10 0.20 MI E OF NEWTON C/L	ST-010-37-01
12593	ON I-65 100 FT S OF NEWTON C/L	INTST-065-37-01
13193	ON I 74 100' E OF BR 4938 IR 19/WESLY RD	INTST-074-54-01
13745	ON I 65 4.60 MI N OF BOONE CO LINE	INTST-065-12-01
14921	ON US 31 0.10MI NORTH OF HAMILTON CO LN	US-031-80-01
15153	ON SR 25 0.10 MI N OF CARROLL CO LN	ST-025-09-01
16841	ON SR 25 1 MI N OF SR 114	ST-025-25-01
17337	ON US 30 0.88 MI WEST OF SR 331	US-030-50-01
18857	ON US 31 0.60 MI S OF SR 250	US-031-36-01
19345	ON I-65 1.00 MI SOUTH OF SHELBY CO. LINE.	INTST-065-03-01
19753	ON US 40 0.10 MI W OF HENERY CO. LINE	US-040-30-01
19921	ON US 50 0.10 MI E OF JENNINGS CO LINE	US-050-69-01
20025	ON I 74 0.10 MI WEST OF FRANKLIN CO LINE	INTST-074-16-01
20073	ON SR 62 0.10 MI EAST OF RIPLEY CO LINE	ST-062-15-01
20377	ON US 52 5 MI W OF FRANKLIN CO LINE	US-052-70-01
22321	ON SR 124 0.20 MI W OF HUNTING C/L	ST-124-85-01
23065	ON I 70 0.10 MI E OF CENTERVILLE RD	INTST-070-89-01
23185	ON SR 28 3 MI E OF US 27	ST-028-68-01
23209	ON SR 26 0.10 M W OF BLACKFORD CO. LN	ST-026-27-01
23417	ON SR 218 0.10 MI WEST OF WELLS CO LINE	ST-218-35-01
23777	ON US 224 0.10 MI E OF SR 1	US-224-90-01
24745	ON SR 120 0.10 MI E OF SR 5	ST-120-44-01
25329	ON US 6 0.10 MI W OF SR 9	US-006-57-01
25641	ON I 69 1.11 MI N OF DEKALB CO LINE	INTST-069-76-01
26001	ON SR 1 0.42 MI NORTH OF US 6	ST-001-17-01
26025	ON US 6 0.10 MI WEST OF OHIO STATE LINE	US-006-17-01
26057	ON US 20 0.10 MI WEST OF OHIO STATE LINE	US-020-76-01
26065	ON SR 120 0.10 MI W OF MICHIGAN STATE LI	ST-120-76-01

Revised Counts Comparisons: Manufactured Goods

Table 4.7 gives the essential statistical results from the comparison of the assigned total manufactured goods flows with each of the truck types included in the road count data. The overall model "explains" about 48% of the variation in total commercial traffic using the flows assigned here to the 40 rural locations. This is a very significant result. Nevertheless, a higher level of explained variation would have been desirable for the overall relationship examined here.

Models derived for the nine types of motor carriers included here were in some cases better and in some cases worst than the overall relationship above. They range from extremely low level relationships for four axle trucks that would be used for deliveries to retail outlets and home deliveries of large retail items (e.g., furniture or appliances), to high relationships for four and five axle trucks with trailers, the most common types of trucks for the movement of manufactured goods. There are some higher and some lower relationships, but these are for motor carriers encountered less often, e.g., some multitrailer vehicles.

This variation in the relationship of the manufactured commodity flows to the road counts for different types of vehicles is not that surprising. Table 4.8 presents the intercorrelations between motor carriers of different sizes and axle configuration. In effect, what the table illustrates is that there are significant differences in these interrelationships and it would be very unlikely for any variable to do a good job at estimating all of these vehicle types.

Revised Counts Comparisons: Total Goods

A further evaluation of the traffic assignments was undertaken using the assignments of all goods (not just manufactured goods) to the Indiana highway network. The overall relationship dropped slightly yielding an adjusted coefficient of determination of .435. This gives an F statistic of 31.07, which is significant at the .001 level. A scattergram of the relationship appears as Figure 4.5.

The Highway Assignments

Given the general acceptability of the cost metric and the assignment obtained for total highway traffic, the next step was to undertake these assignments for each of the manufactured commodities of interest here. This was done and some of major maps resulting for the 1993 flows are displayed in Figure 4.6 through Figure 4.11. Unfortunately, there is no way to check the accuracy of the individual assignments and one is left with accepting or rejecting them based on their appearance. Before doing the latter the reader should bear in mind the statistical tests that were undertaken to verify the assignment procedure.

Table 4.7. Comparison of Assigned Flows with Truck Flows at 40 Selected Locations

Variable	Mean	Stand. Dev.	Coefficient	Intercept	F	R^2
Total Commercial Flow	2635	3525	1.27	1273	34.8	.49
Two axle Trucks	198	183	.054	139	18.6	.31
Three axle Trucks	102	113	.036	63	23.0	.38
Four axle Trucks	19	20.5	.004	15	5.3	.12
Four axle Trucks w/trailer	259	320	.137	111	78.4	.67
Five axle Trucks w/trailer	1671	2429	.977	624	56.0	.60
Six axle Trucks w/trailer	69	117	.022	45	5.65	.13
Five axle Trucks with Multi trailer	292	1416	.021	269	.032	.00
Six axle Trucks with Multi trailer	13	24	.011	1	97.9	.72
Seven axle Trucks with Multi trailer	13	41	.008	4.2	6.05	.14

Table 4.8. Intercorrelations of the Motor Carrier Types Based on 2417 Locations

	Vt-5	Vt-6	Vt-7	Vt-8	Vt-9	Vt-10	Vt-11	Vt-12	Vt-13	Tot Com Veh
Vt-5	1.00	.58	.49	.65	.62	.40	.24	.48	.15	.71
Vt-6	.58	1.00	.62	.64	.59	.46	.20	.48	.15	.67
Vt-7	.49	.62	1.00	.45	.42	.31	.14	.29	.06	.48
Vt-8	.65	.64	.45	1.00	.83	.58	.33	.73	.16	.89
Vt-9	.62	.59	.42	.83	1.00	.65	.37	.83	.19	.98
Vt-10	.40	.46	.31	.58	.65	1.00	.23	.61	.39	.68
Vt-11	.24	.20	.14	.33	.37	.23	1.00	.37	.09	.44
Vt-12	.48	.48	.29	.73	.83	.61	.37	1.00	.24	.83
Vt-13	.15	.15	.06	.16	.19	.39	.09	.24	1.00	.25
Tot Com Veh	.71	.67	.48	.89	.98	.68	.44	.83	.25	1.00

Commercial Vehicle Counts vs. Modeled Industrial Flows

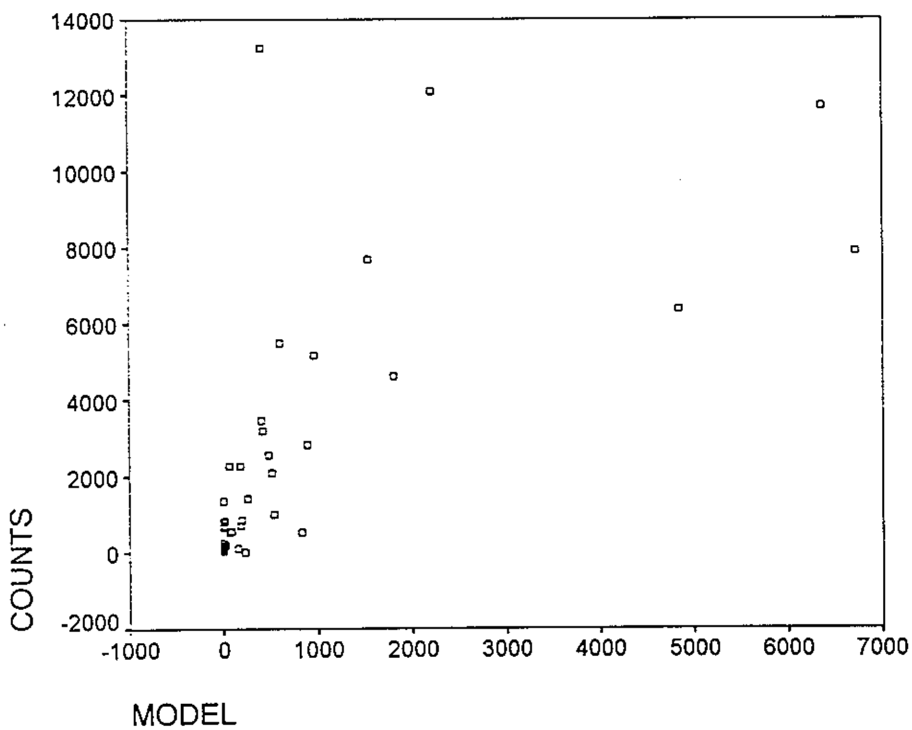

Figure 4.5 Scatter Diagram of Relationship of Model Estimates and Counts

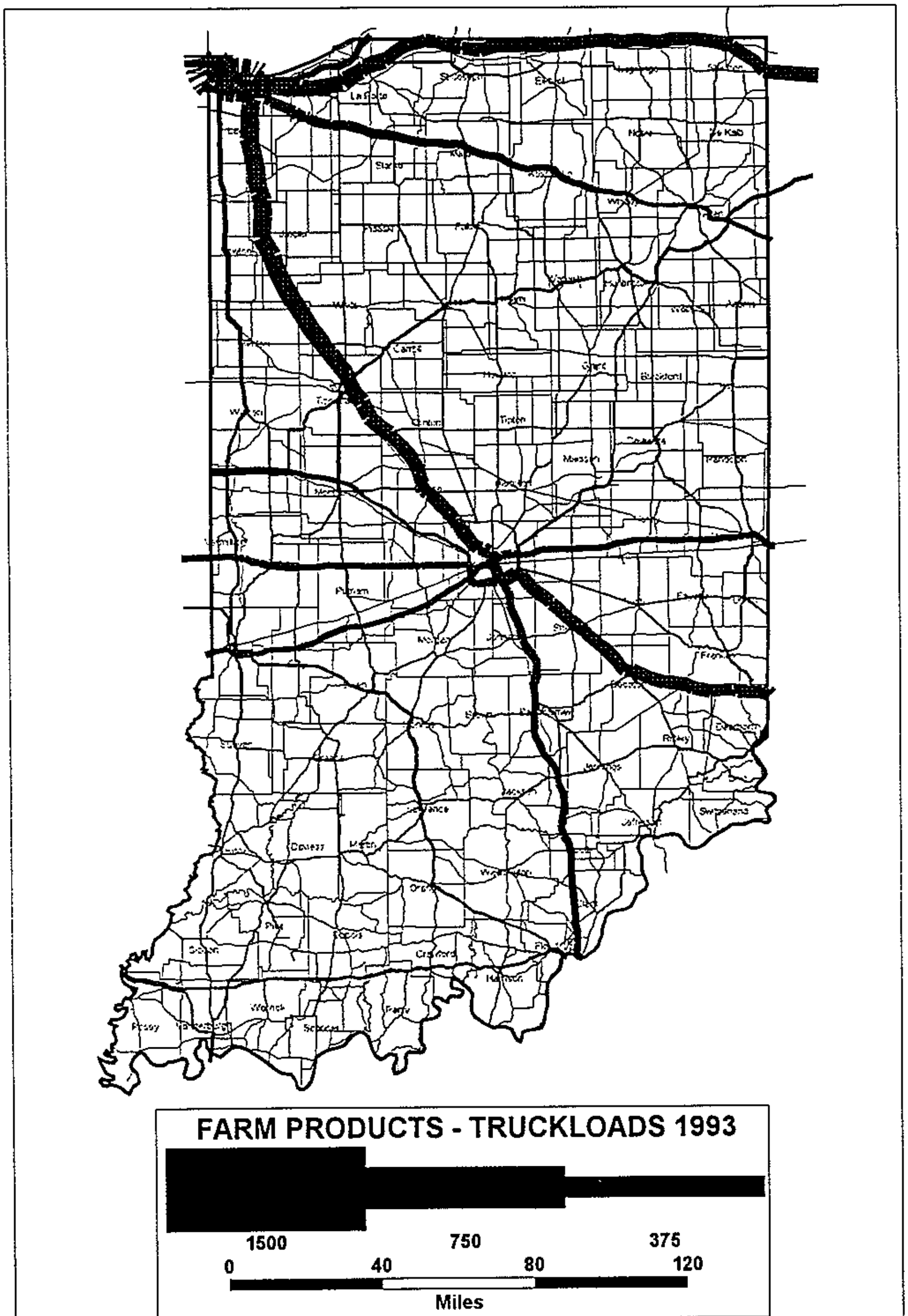

Figure 4.6 Daily Motor Carrier Volumes - Farm Products 1993

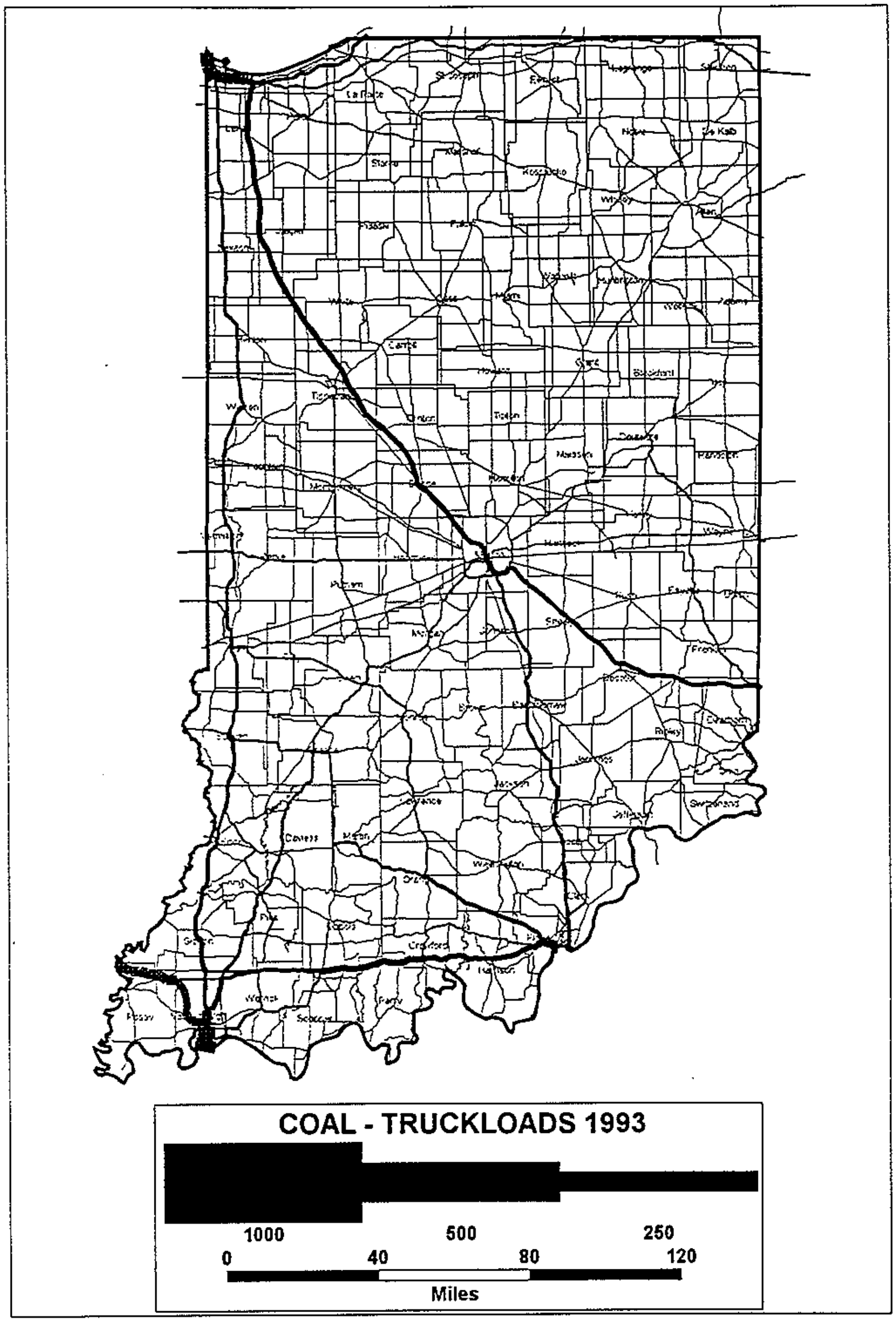

Figure 4.7 Daily Motor Carrier Volumes - Coal 1993

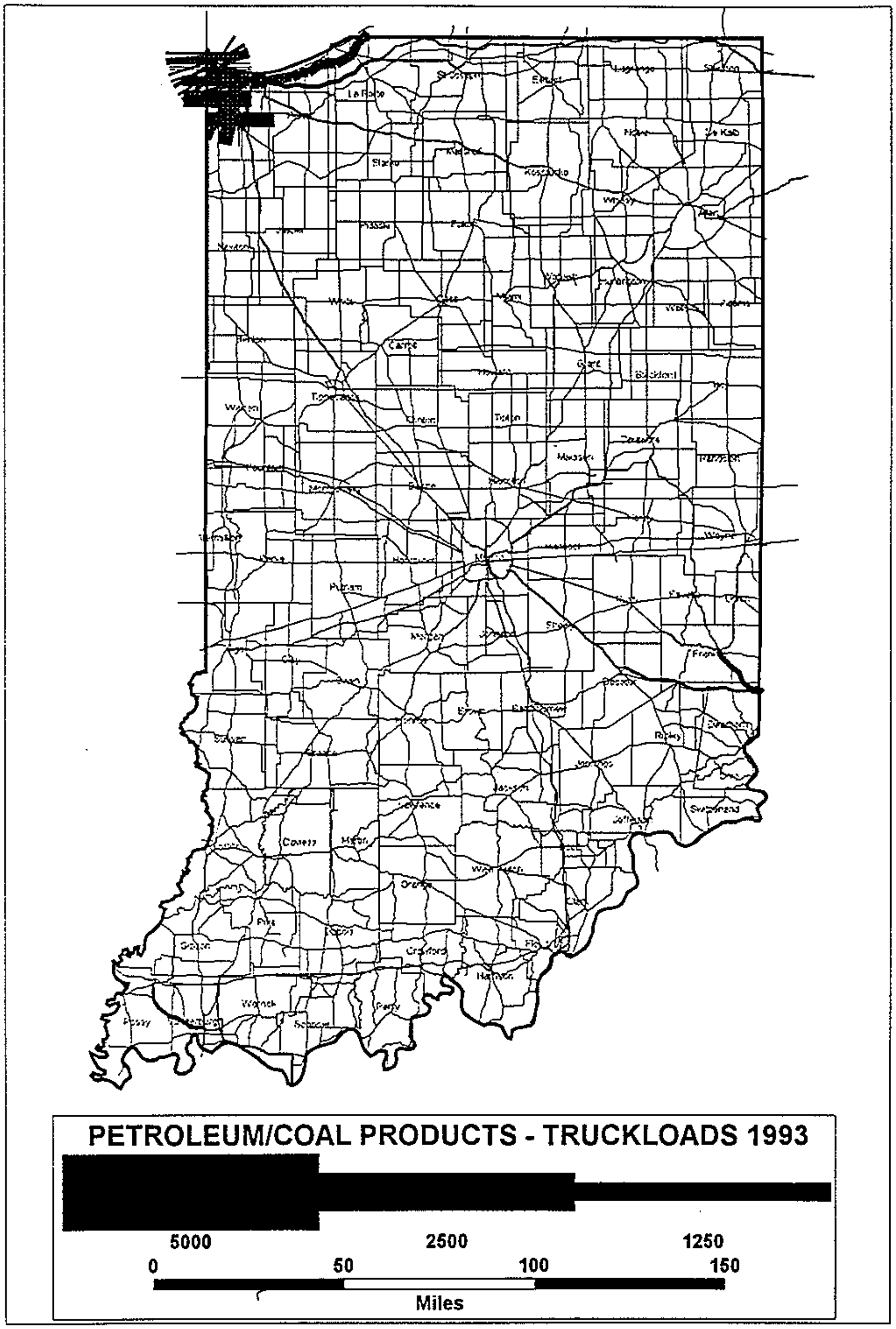

Figure 4.8 Daily Motor Carrier Volumes - Petroleum and Coal Products 1993

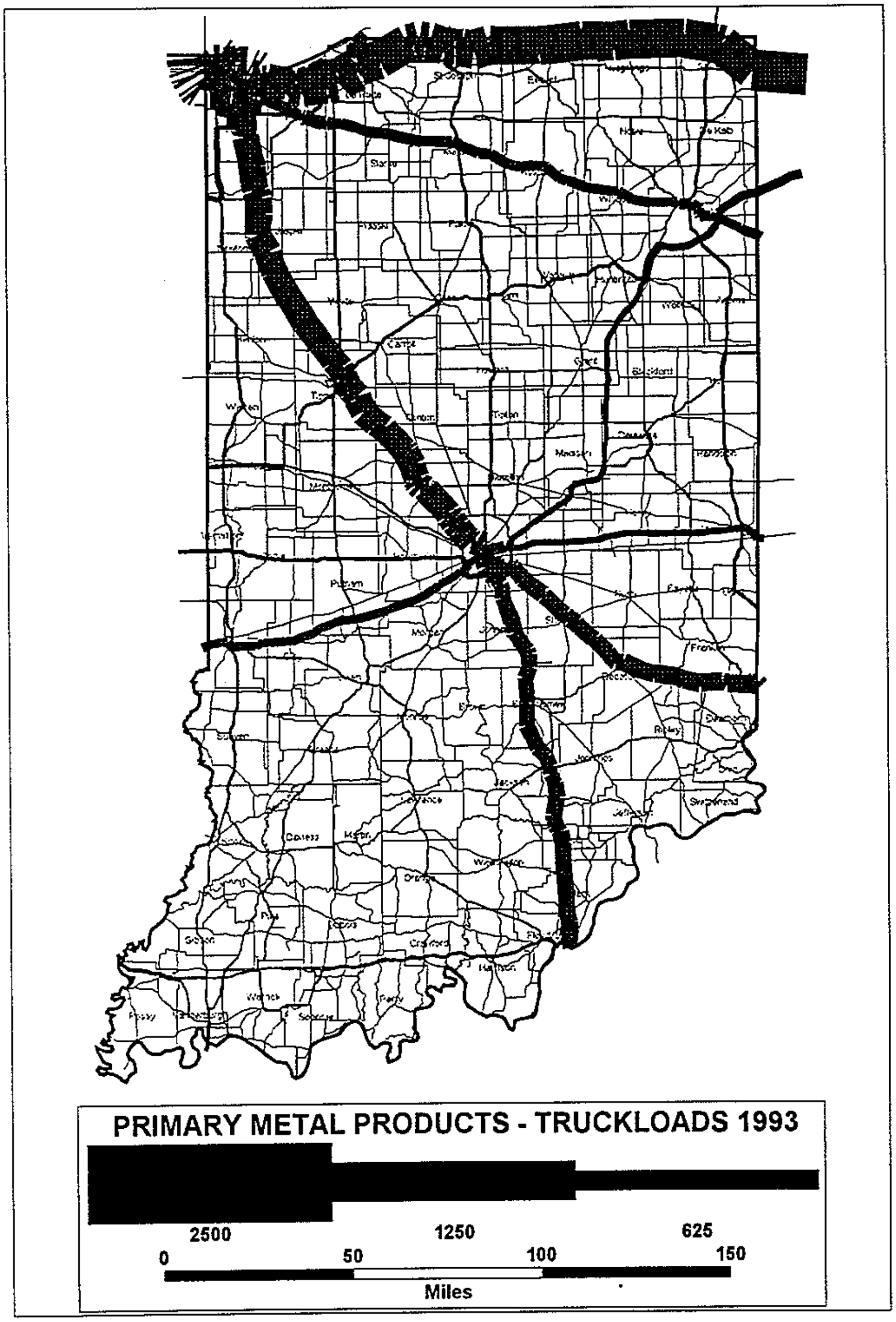

Figure 4.9 Daily Motor Carrier Volumes - Primary Metal Products 1993

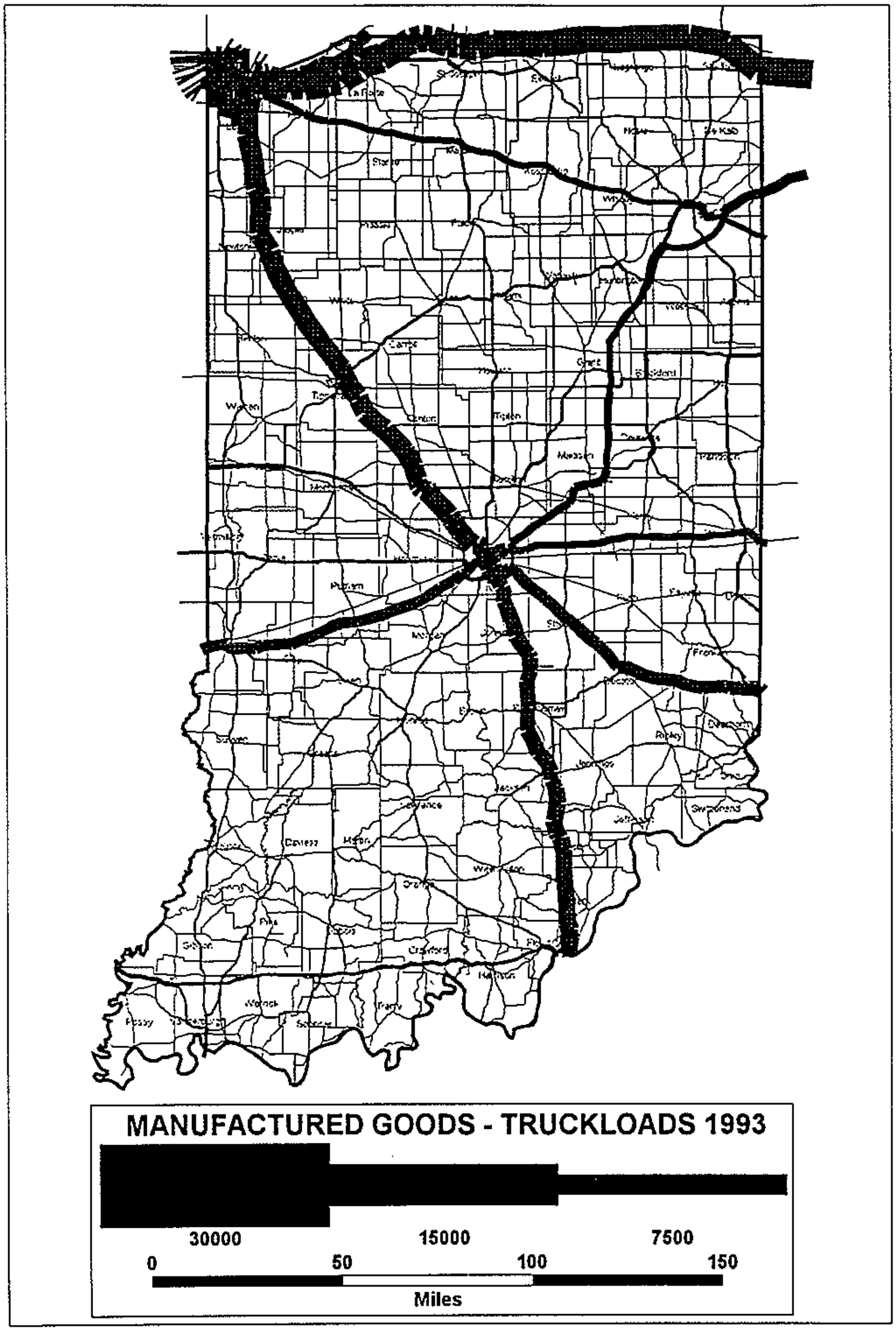

Figure 4.10 Daily Motor Carrier Volumes - Manufactured Goods 1993

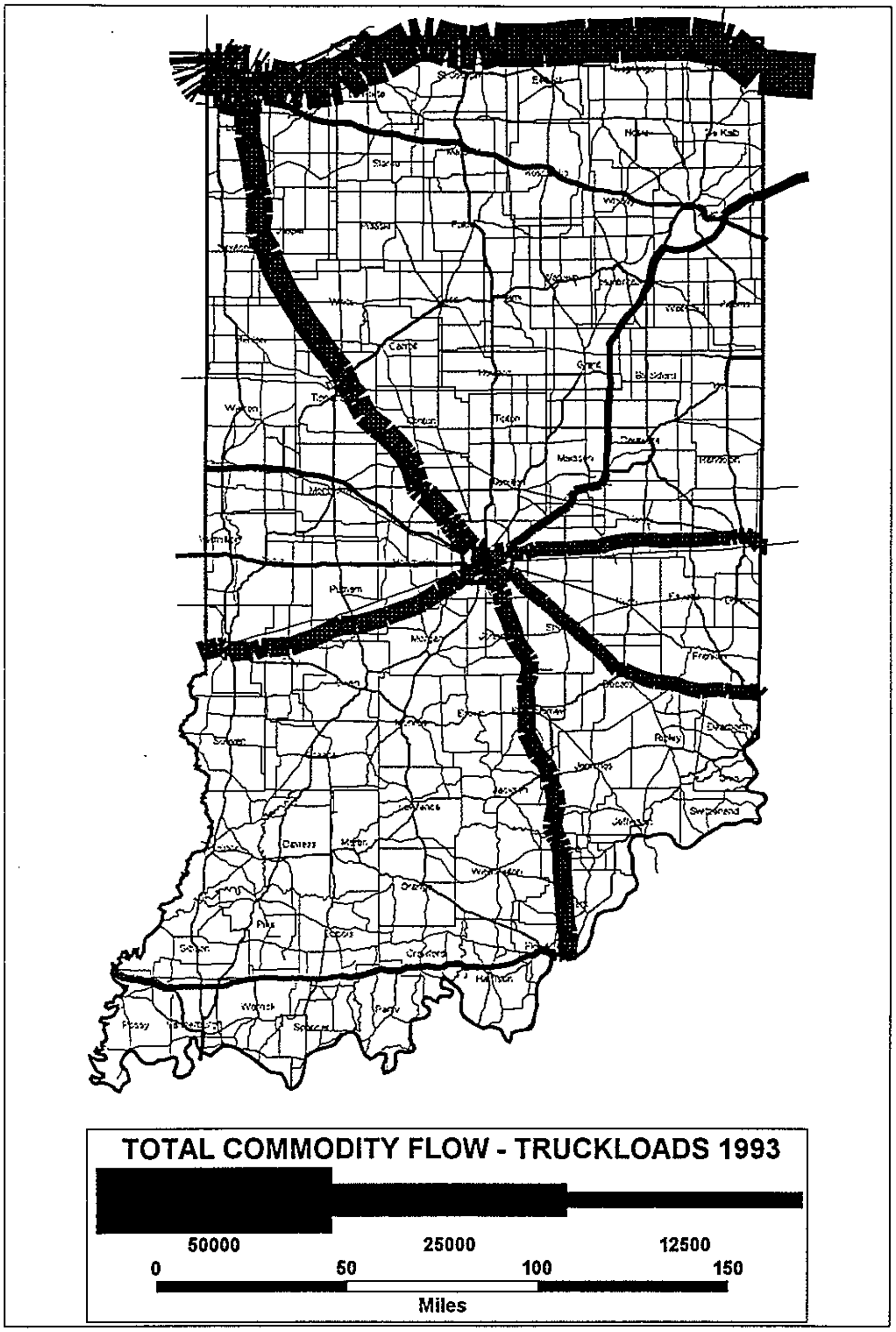

Figure 4.11 Daily Motor Carrier Volumes - Total Traffic 1993

Sources of Errors in the Assignments

It is reasonable to offer some explanation for the errors observed in the assignment process, but in order to do this it is also necessary to examine possible sources of error throughout the planning and analysis effort undertaken here. Some of these reasons have been stated previously, but they are restated here simply as a catalog of items that should be considered in future applications.

Network and Nodal Definition

The network used in this study is an abbreviated representation of the state highway network. In other words there are places in Indiana where a substantial amount of travel occurs on local or county roads. This study includes no county roads and not all highways or major urban streets were included. These factors suggest that higher volumes may have been assigned to some of the roads included here, i.e., the roads where the traffic actually occurs are not included here.

A second network related problem resulted in some highways of the state simply ending at the state border. Although all of the FHWA digital highway planning network contiguous to the state's border were included in the analysis, the network within Indiana was much more detailed than the FHWA network. This resulted in some Indiana highways ending at the border. The problem could have been remedied by digitizing new links into the FHWA network, but this was beyond the scope of this project.

Still another source of error related to the network is the placement of external nodes. More specifically, this study assigned all traffic in the states and counties to a point or centroid, although not necessarily the geometrical center of these areas. States bordering Indiana were given additional nodes under the premise that it was desirable to allocate traffic to more than a single centroid for states very close to Indiana. For example, allocating all of the traffic generated by Ohio to Columbus, Ohio, would not represent the Toledo-Cleveland area of the north or the Cincinnati area in the south. Some research is necessary on the placement of these nodes, since it can clearly impact the results here. Unfortunately, a database that would enable planners to research this problem does not exist, unless the 1993 Commodity Flow Survey could be modified to answer this question

A related point is that beyond the circular highway network that surrounds Indiana all of the highways are part of the Interstate Highway System. It was believed that the circular network would eliminate bias in the traffic assignments, but the impact of all the traffic coming into this region on links of the Interstate Highway System may make it more attractive to simply stay on these routes. This would also result in less traffic finding its way to lesser routes. In addition, the use of single centroids for states and the use of the Interstate Highway System may cause another source of bias in that traffic may be "loaded" onto the network in the wrong location, e.g., all traffic from California is loaded onto the Interstate System near San Francisco, as opposed to a node in Southern California and this results in southern Interstate Highways (Interstates 40, 10 or 20) being used less by the assignment process.

Traffic Generation: Production and Attraction

This study made use of the 1993 Commodity Flow Survey. Numerous sources of error attributable to traffic generation models were avoided. In particular the data used in the traffic production portion of the study used those data at the state level in virtually the same form in which they were published. Errors in this section would be the errors of the sampling and data expansion process used. Models were used to distribute the manufacturing flows produced or attracted by Indiana's counties, but the aggregate amount of these were constrained to the totals derived from the 1993 Commodity Flow Survey. Nevertheless, the models used for Indiana traffic are not perfect, and this could also result in some error in the estimates of production and attraction.

All manufacturing traffic produced in the United States is included in this study. Even traffic that is produced in the U.S. and destined for Japan or another Asian country via Los Angeles or another port in that state is included here as a shipment to California. On the other hand manufactured goods shipped to the U.S. from a foreign country is missing in the data used here, to some extent. It was noted previously that the value of all imports was assumed to be equal to the value of all exports. This results in slightly larger allocations of traffic being attributed to some states. Nevertheless, since the U.S. tends to import more manufactured goods than it exports, this traffic is lost to the modeling since the origin and destination of such flows are not generally available. These data are collected for some port cities and may be purchased, but their costs were beyond the resources allocated to this project.

Traffic Distribution

The fully-constrained gravity model used here yields flow estimates that are consistent with the initial inputs to the modeling process. Traffic produced and attracted according to the model are equal to the actual values used in the modeling. In addition, the average shipment length observed tends to be replicated (in most cases) by the modeling process.

This project did not perfectly replicate the average shipping distance. The major reason for this is the average shipping distances between states in the western U.S. It is not possible to constrain flows if distances between places are very large, relative to observed averages. In some cases this was not a critical point, but in others the average value could not be constrained. Even in those cases where the flow distances are constrained to inputted values, it is possible that the actual flows may differ from those generated by the model. This is not a likely outcome, but it is a possible outcome.

This study actually went further than any other study completed to date in that it examined the average length of shipments originating for a subarea. This was Indiana originated flows in this case. The modeling was in no way constrained to replicate Indiana's average length of shipments, but this piece of information was generated and compared with data from the 1993 Commodity Flow Survey used here. In some cases errors were permitted in the U.S. flow portion of the study so that the Indiana average could be closer to the observed average value for this.

In effect, the distribution modeling does not seem to be a major source of error here. It may be one of the more accurate portions of the study.

Modal Split of the Traffic

The NEWMODE modal split program developed for this project assigns traffic to a variety of modes based on observed patterns in the 1993 Commodity Flow Survey. The major modes of interest here are highways and railroads. Some of the mode choices available involved two modes, e.g., rail and highway, or air transport and highway. These were assigned to both of the principle modes (rail or highway), if appropriate. This results in slightly higher total flows, but this difference is not very significant as the reader can see by examining the modal assignment probabilities in the appendix.

All of the modal allocations were made in tons. After these allocations the traffic is divided by density factors consistent with specific classes of manufactured goods. In some cases the density factors (tons per vehicle unit) are quite different for goods received by a state and for goods shipped by that same state. For example, there are significant differences in the weights of television components and completed television sets, but these are both in the same manufactured goods STCC group. Weighted average density values were used here, but these could result in more or less vehicle units depending on whether the density values are too low or too high.

Aside from these points this does not appear to be a major source of potential error in the modeling undertaken here. While errors are possible here they appear to have only minor impacts on the model outputs.

Traffic Assignment

There was a substantial discussion of traffic assignment earlier in this chapter. The reader was informed of the different assignment routines that could have been used, and the decision to use an "all or nothing" assignment procedure here. This method is not the most popular because it fails to consider link capacity or the consequences of congestion on route choice by highway users. These are certainly valid criticisms for the urban application of these methods, but in large scale regional studies covering multi-state areas it seems an appropriate technique since congestion and capacity are of less concern at that scale, particularly in the Midwest.

There are an infinite number of ways to measure the cost of transport, it is not possible to say that the method selected here is as good as another not evaluated [8]. There is no indication that the cost selected resulted in any errors.

It might be viewed as desirable to add capacities to all the links examined here. However, none of the state's flows were close to their known capacity. Also, these values were not known for the highways outside of Indiana.

It actually seems that traffic flows at the scale examined here are in need of a new traffic assignment method. This method would look at the three or four best (e.g., lowest cost) routes that could be taken between an origin and a destination and assign probabilities to these. Trucks would be assigned in a Monte Carlo fashion with assignments proportional to their probabilities. For example, truckers passing through Chicago en route to Pittsburgh may consider Interstates 80, 65-70, or even U.S. 30. The proposed method would consider all of these as possible choices. This would overcome some of the shortcomings of the single least cost route of "all or nothing" assignments being selected all the time.

The Railway Network and the Cost of Movement

It may appear anticlimactic, but in this section we will examine the assignment of traffic to the railway network. Many of the points made in the previous section with regard to the highway traffic assignments also apply here; this is particularly true with regard to the sources of possible errors in the assignment process. At the same time the entire rail traffic assignment process is different enough that it merits its own discussion.

As noted in Chapter 2 the digital rail network used here is the 1:2,000,000 rail network prepared by the Federal Railroad Administration. Although other digital representations of the U.S. rail network exist, these generally lack attribute data that are necessary for use of the networks in traffic assignment. It is for this reason that this older network, dating from 1992, was used here.

Although highway traffic assignment is controlled by travel time and the user's desire to minimize this, or a cost version of the travel time, railway operations are not so preoccupied with this. It is certainly true that railroads want to move from origin to destination quickly, but speed is usually measured in days for railroads as compared to hours for highways. However, the problem is actually more complex than this.

The movement of a primary or manufactured product from an origin to a destination generates a waybill, which is a statement of the charges to the shipper for the move. It is a statement of the total rail transport charges. The actual move may have been made by a single railroad, which was responsible for picking up the rail car at the origin station or siding, moving it over its system, and terminating it, or dropping it off, on the destination side. Or the move could have been made by multiple rail carriers, one on an origin branch line, two railroads that handled the move across the country, and a fourth railroad on some destination branch line. Or perhaps six or seven railroads were involved in the move. All of these possible scenarios create problems for the assignment of traffic to the rail network.

Each of the railroads involved in transporting a rail car of goods and products is interested in maximizing their income. They wish to have the lion's share of the charges on the above mentioned waybill. However, the railroads involved in the move "divide" these charges into what are called "divisions," and this represents the income the railroad makes from the move. In general, the originating railroad may get anywhere from 8% to 20% of the revenue, a terminating railroad would get a similar division, and the carriers responsible for the majority of the miles between the origin branch line and the destination branch line would divide that portion of the revenue remaining based on the miles that each transported the rail car. If the intervening distance is 1,000 miles, and carrier A handled it for 800 miles and carrier B handled it for 200 miles, the two railroads would split the revenue 80% to 20%. It should be apparent that each of the intermediate railroads want to hold onto the traffic as long as possible, in a distance sense, since this would increase their income, other things being equal. Of course there are limits to the total mileage charges, but railroads nevertheless have a lot of freedom in routing traffic.

In this environment it should be apparent that the carriers do not follow a shortest path approach to routing traffic. The logical approach to rail traffic assignment would appear to be something like a shortest path route, with turning penalties. In other words if carrier A picks up traffic it would move it from the origin to the destination in such a manner as to minimize distance traveled, assuming the destination is on its system. If this is not the case the turning penalty (assuming it is set reasonably high) will prevent the traffic from being passed over to a second carrier until the last possible minute. If local or regional railroads were responsible for the originating or terminating traffic, such an assignment process would have little impact on that portion of the move. The rail car must go onto these originating and terminating carriers lines. A series of experiments with shortest paths including turning penalties carried out during an earlier rail planning project suggested that this approach to rail traffic assignments was non-workable [9].[2]

[2]It is quite possible that the failure of this algorithmic approach to the problem - minimal path with turning penalties - failed because of a significant duplication of rail links in the Midwest area. In effect for a large section of the region each rail line had another rail line on top of it connecting the same endpoints, but identified as a different link. The use of turning penalties were useless in that assignment process, and this may have been due to the fact that traffic would be assigned to the underlying link if the top link had a turning penalty. This problem was not recognized in that earlier study, but became apparent during this project.

As a result of the failure of the turning penalties to work, an alternative approach to rail traffic assignment was pursued.

Although there is some desire on the part of rail carriers to minimize the length of haul, they have a tendency to use mainline trackage even though secondary lines may be more direct. The question was how to represent this tendency with the rail data available for the digital network. Track condition plays a part in such decisions, but this is a very dynamic variable that would change more frequently than the database available. It seemed a new measure of spatial separation was necessary. The new measure of spatial separation would still incorporate an attempt to minimize shipping distance, but it would also pick those routes that the railroads tend to use.

Short line or regional railroads that originate or terminate traffic are not important in this methodology, since the origin and destination of shipments must be reached. In other words these moves can be replicated by any methodology regardless of the cost attached to it simply because the endnodes of these moves are used as input to the methods.

The measure finally adopted had the form

$$I = (L\ (1/(D+1)))$$

where I = the index of spatial separation;
L = the length of the line segment of the network; and,
D = the traffic density of the line in millions of gross ton-miles per year.

This measure diminishes the length of line segments by dividing the segment by its traffic density, i.e., by gross ton-miles per year. Typical traffic density values vary from 0 to about six million gross ton-miles per mile of line.

If we have five route segments of 100 miles in length each with traffic density ranging from 0 to 1 to 2 to 3 to 4, the index of spatial separation would be 100, 50, 33, 25, and 20. When used on lines with high traffic density these routes "become shorter" and are always selected. Lines of low traffic density, do not become "longer" since their traffic density always has a unit value added to it. Lines of 0 traffic density would become lines of 0 length, if it were not for this correction factor.

The transport cost matrix used for assigning rail traffic was defined using the length-density index described above. Although it is beyond the scope of the present study, it would be desirable to have a study undertaken that would evaluate a broad array of indices (including the one utilized here) and methods of assigning traffic to a rail network. Such a study would require the existence of a set of actual flows, referred to in the highway case as target flows, but these are generally not available in the rail case.

Target Flows

Target flows in the highway case are actually road counts of vehicles. In some cases the vehicles are broken into groups, e.g., commercial vehicles, and these may be used as a variable that assigned flows should resemble. In the railway case there are no target values that are route segment specific. Data that are made available in the public use carload waybill sample are too gross to be used for this purpose. Very detailed information that would allow comparisons are available in a complicated fashion for flows involving the movement to, from, or through the state of Indiana, but translating compiled data into this format is difficult.[3] As a result one must visually examine the flows to see if they are consistent with expectations.

Flows Assigned

The primary and manufactured commodities assigned to the rail network do not come from the carload waybill sample mentioned previously, but are a product of this project and programs developed by it. More specifically, the traffic assigned is the product of NEWMODE.EXE, a computer program that splits commodity specific traffic between modes based on the length of haul. The basis for these splits is data published for the United States in the 1993 Commodity Flow Survey. In the case of some manufactured goods this is a very small amount of traffic since railroads during the latter part of the 20th century have loss significant market share for many manufactured goods to motor carriers.

The graphic results of the traffic assignment process for total rail traffic appears as Figure 4.12. Traffic assignments for five other major commodity groupings appear as Figures 4.13 to 4.17.

Sources of Error

The sources of error are in many cases the same as they are in the highway case. These include problems in defining the network and the nodes on it, problems in estimating traffic produced and attracted, problems with the distribution model used or the modal split procedures, or simply problems with the assignment process used to represent the routes taken by the traffic examined here. In the best of all worlds these sources of error would be eliminated, but this would require access to proprietary railroad data that the industry is unwilling to release due to the potential negative impacts this could have on competition and modal share. In the case of railroads this is not an irrelevant point since they are no longer the dominant mode for any manufactured goods transported today and in some cases account for a very small fraction of the traffic moved.

---

[3] It has come to our attention that Caliper Corporation, Inc., the developers of TransCAD are currently working on this problem for the Federal Railroad Administration of the U.S. Department of Transportation.

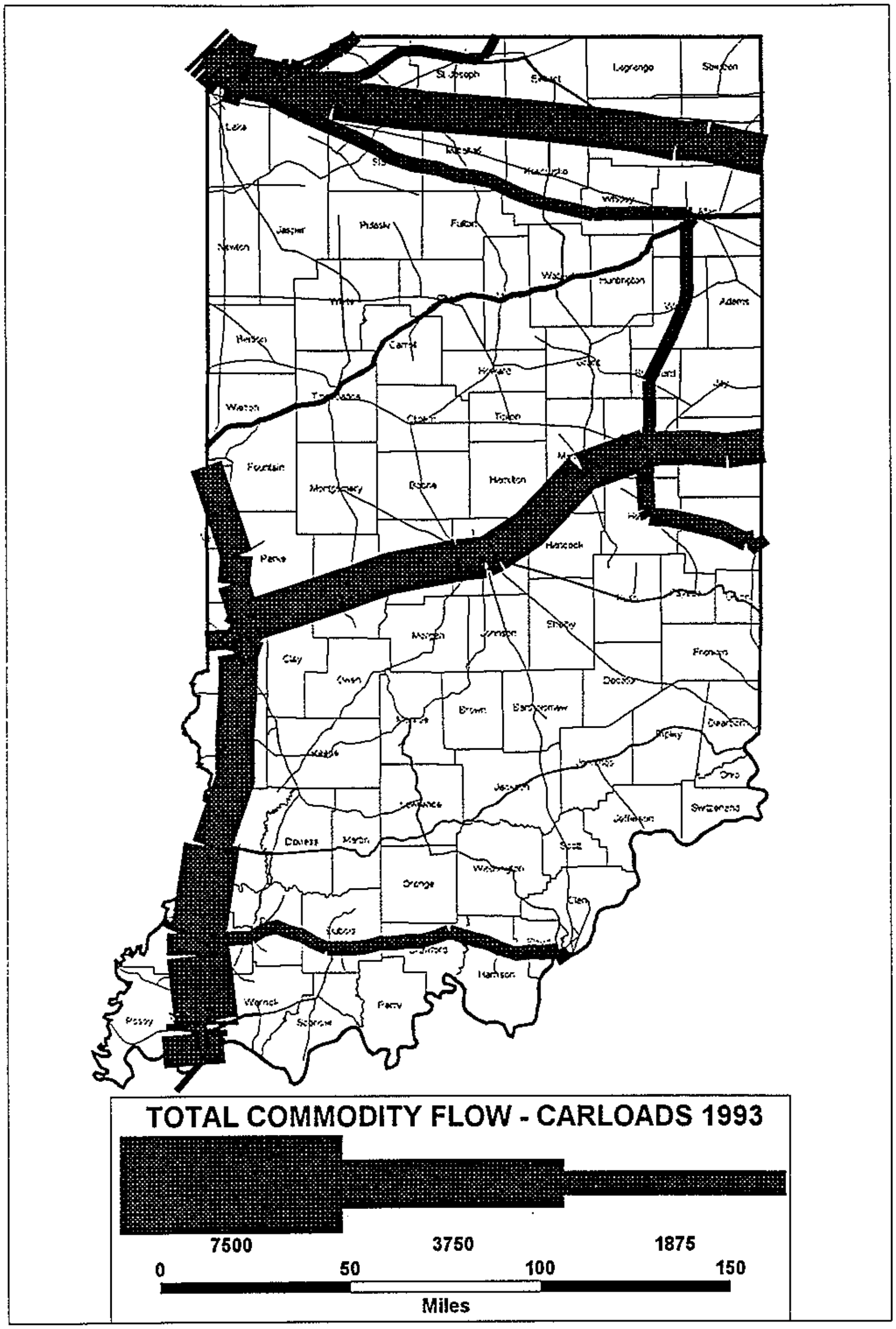

Figure 4.12 Total Daily Rail Traffic (Carloads) 1993

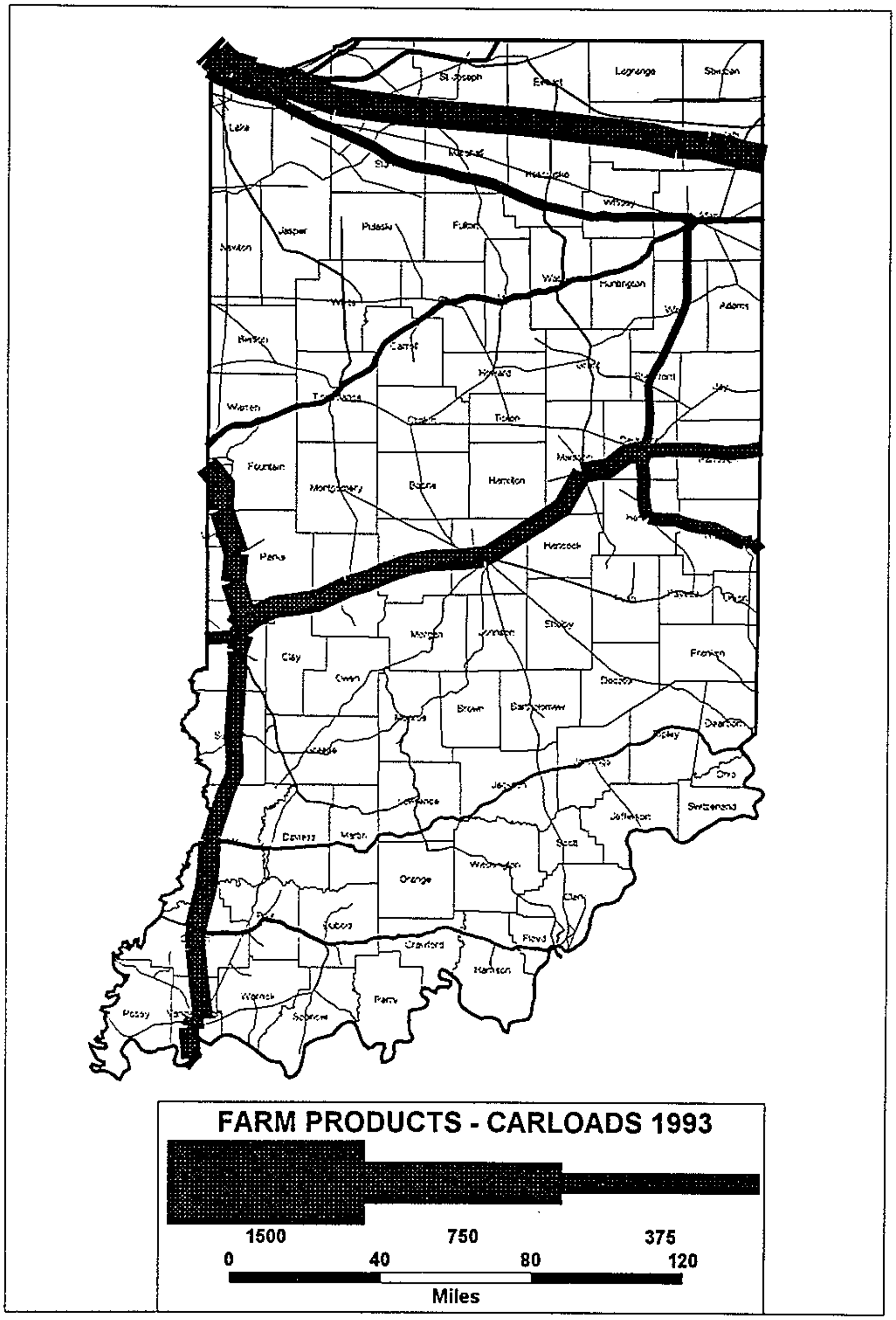

Figure 4.13 Daily Railroad Carloads - Farm Products 1993

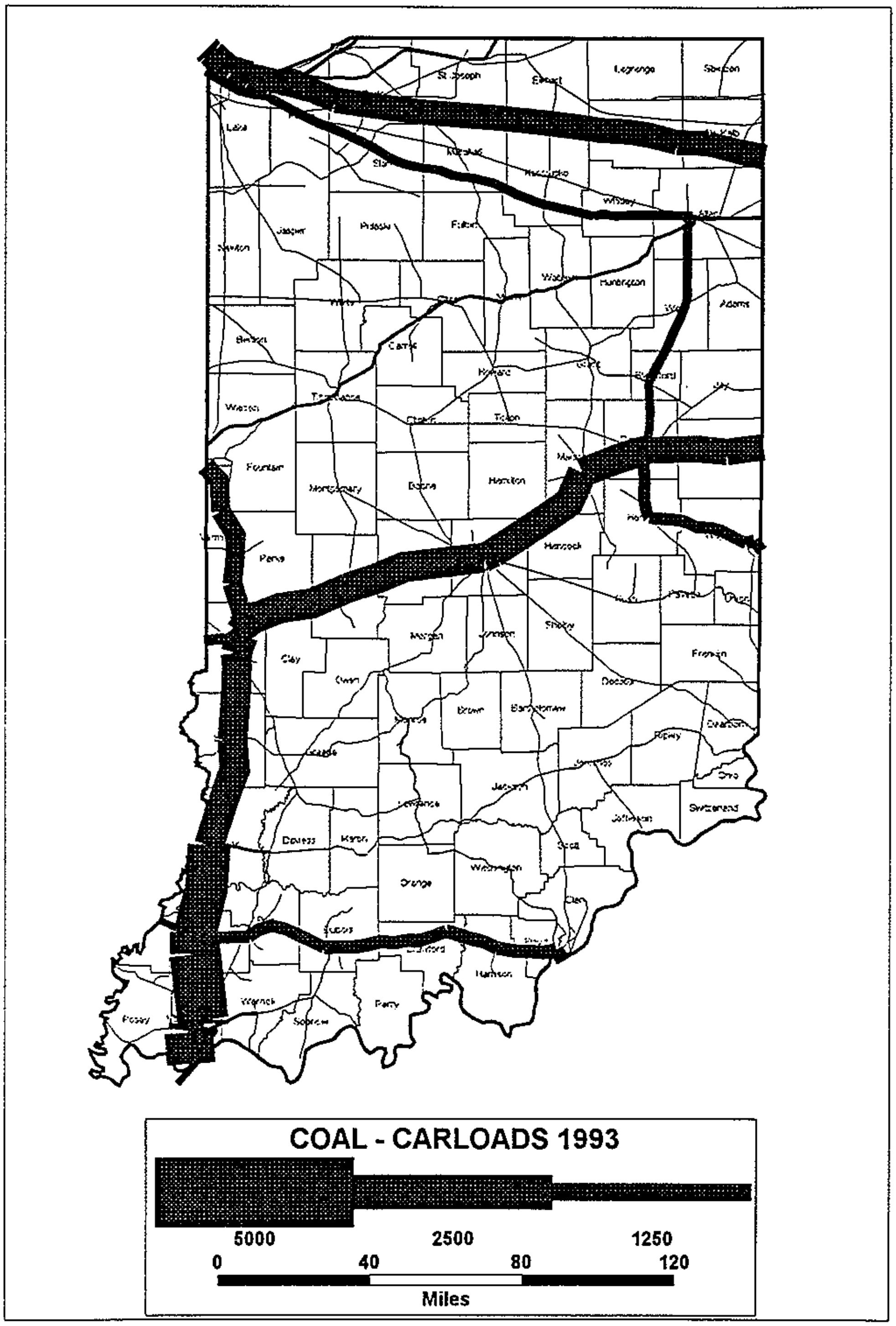

Figure 4.14 Daily Railroad Carloads - Coal 1993

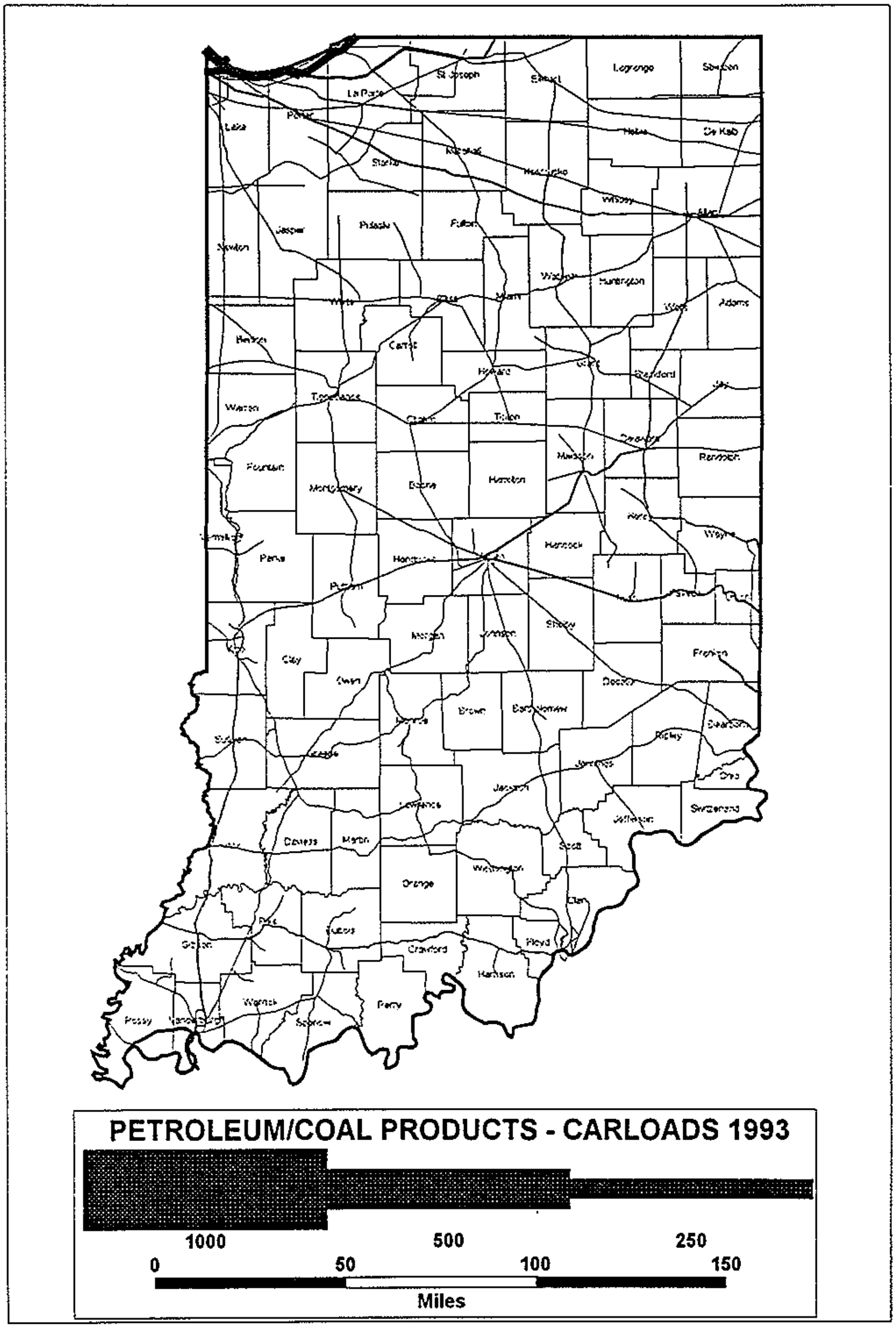

Figure 4.15 Daily Railroad Carloads - Petroleum and Coal Products 1993

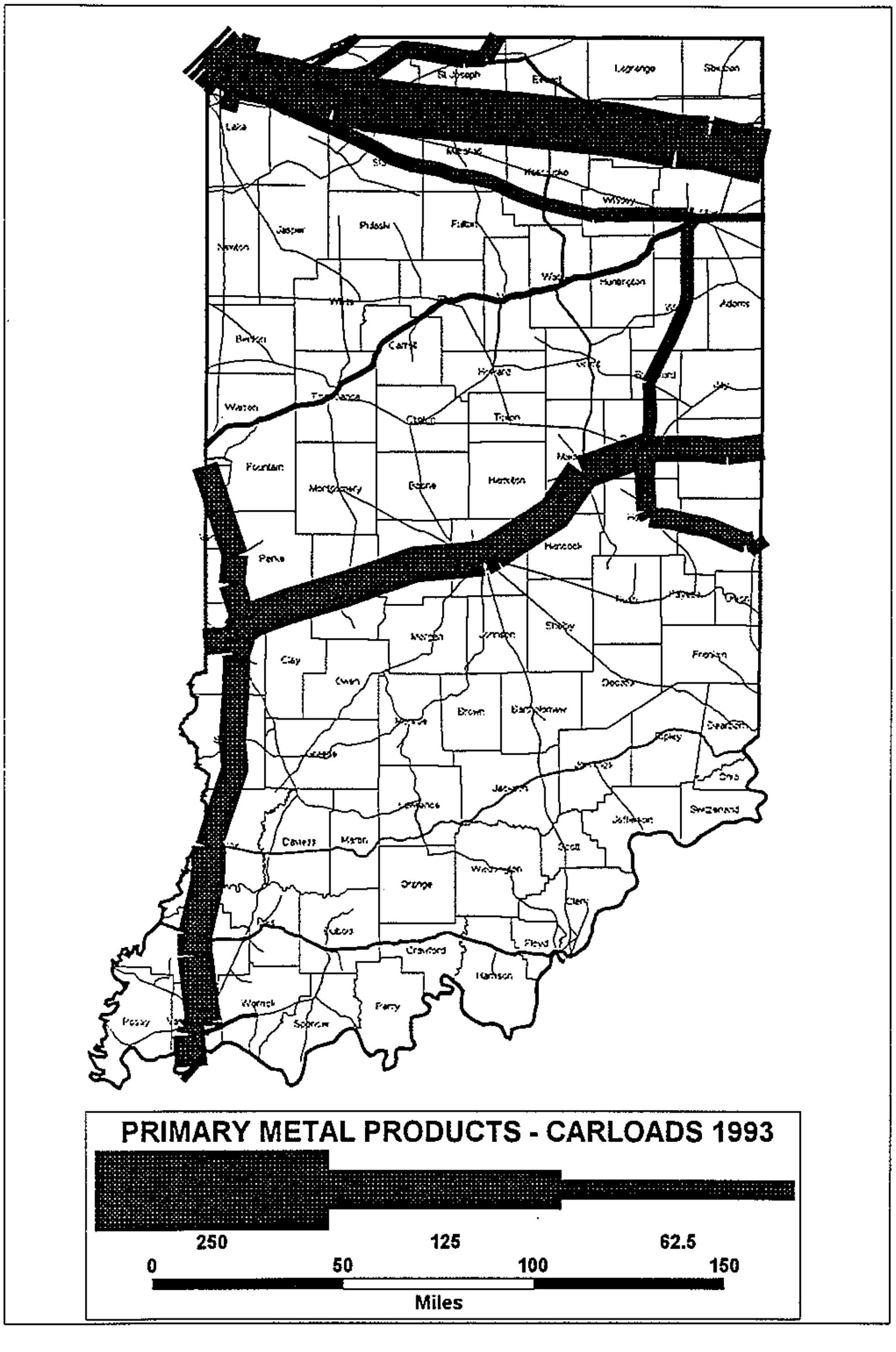

Figure 4.16 Daily Railroad Carloads - Primary Metal Products 1993

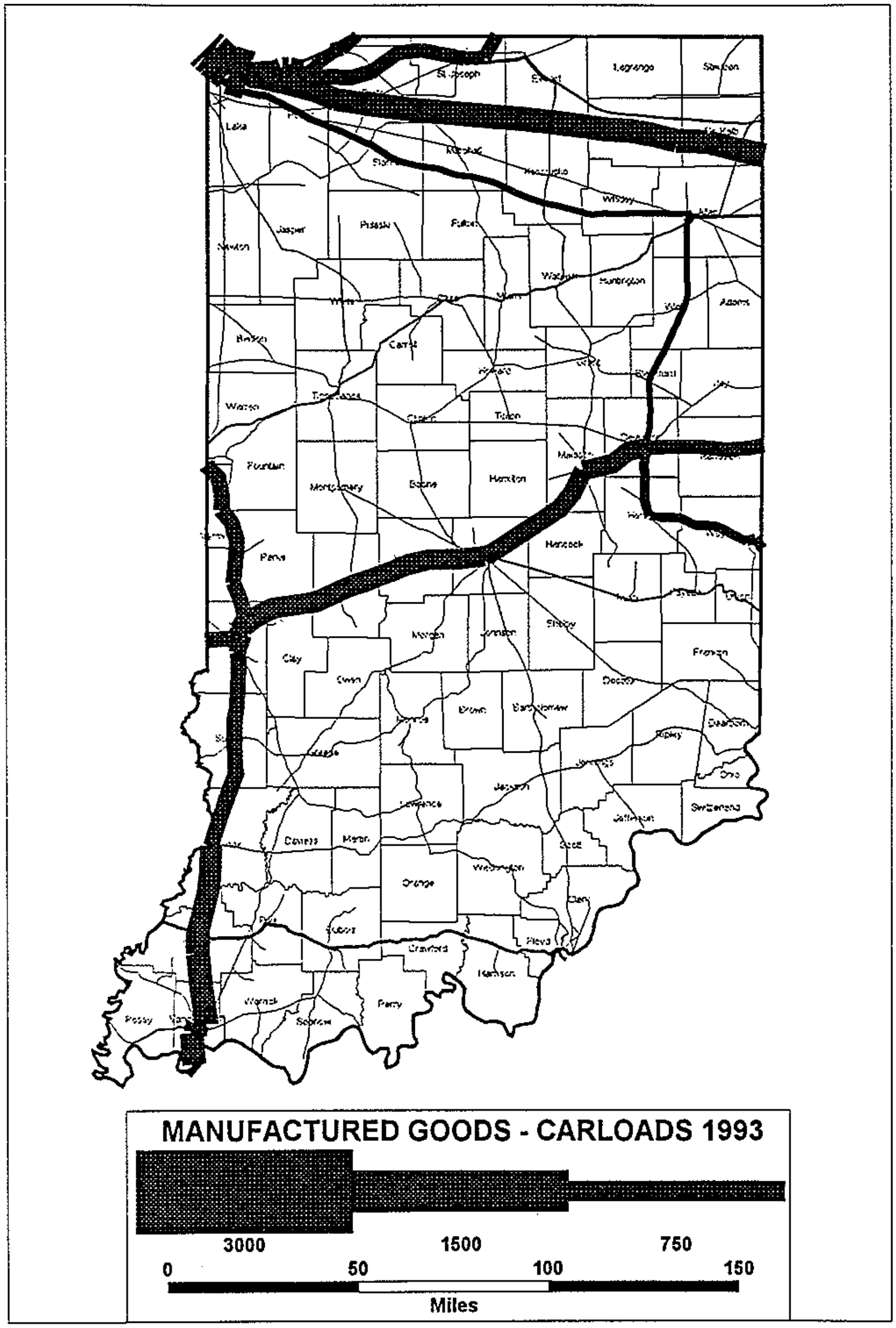

Figure 4.17 Daily Railroad Carloads - Manufactured Goods 1993

Conclusion

In the following chapter we will examine the traffic forecasts for 2005 and 2015. It is easy to lose sight of the fact that the primary objective of this entire chapter has been to develop estimates of traffic that can be projected into the future. It would appear that the models developed and the methods used can replicate existing flows and presumably these will be equally accurate for future flows if the parameters derived remain constant. This is the fundamental premise underlying all future-oriented transport planning and analysis; it is assumed that such stability exists.

References Cited

[1] Bureau of the Census (1996), *1992 Census of Transportation, Communications, and Utilities, "1993 Commodity Flow Survey,"* Washington, D.C.: U.S. Department of Commerce.

[2] Bureau of the Census (1982), *1977 Census of Transportation*, Washington, D.C.: U.S. Department of Commerce.

[3] Bureau of the Census (1979), *County Business Patterns*, Washington, D.C.: U.S. Department of Commerce.

[4] Wilson, A.G. (1967), A Statistical Theory of Spatial Distribution Models. *Transportation Research 1 (3)*, pp. 252-270.

[5] Wilson, A.G. (1970), *Entropy in Urban and Regional Modelling,* London: Pion.

[6] Black, W.R. (1995), Spatial Interaction Modelling Using Artificial Neural Networks, *Journal of Transport Geography 3 (3)*, pp. 159-166.

[7] Black, W.R. and Larson, R.C. (1972), A Comparative Evaluation of Alternate Friction Factors in the Gravity Model. *The Professional Geographer 24 (4)*, pp. 335-337.

[8] Transportation Research Board (1992), *Highway Capacity Manual, 2nd Edition, Revised, Special Report 209*, Washington, D.C.: National Research Council.

[9] Black, W.R., Palmer, J.A., Munn, D., Martin, R. and Black, R.J. (1995), *A Preliminary 1995 Indiana State Rail Plan,* Bloomington, IN: Department of Geography and the Transportation Research Center, Indiana University.